UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

INFERENCIA ESTADÍSTICA APLICADA EN LA GENERACIÓN DE UNA PROPUESTA DE HORARIOS PARA LAS CARRERAS DEL DEPARTAMENTO DE MATEMÁTICAS

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIA

PRESENTA:
MIRIAM GABRIELA COLÍN NÚÑEZ

TUTOR
DR. ARRIGO COEN CORIA

CIUDAD UNIVERSITARIA, CD. MX., 2021

Datos de la Alumna:

Colín

Núñez

Miriam Gabriela

(Teléfono)

Universidad Nacional Autónoma de México

Facultad de Ciencias

Actuaría

(número de cuenta)

Datos del tutor:

Dr.

Arrigo

Coen

Coria

Datos del sinodal 1:

Datos del sinodal 2:

Datos del sinodal 3:

Datos del sinodal 4:

Datos del sinodal 5:

Datos del trabajo escrito:

Inferencia estadística aplicada en la generación de una propuesta de horarios para las carreras del departamento de matemáticas

(Número de Páginas)

2021

Dedicado a

Agradecimientos

¡Muchas gracias a todos!

Índice general

Índice de figuras

Índice de tablas

Códigos

Capítulo 1

Introducción

En este trabajo se hará un análisis estadístico de los datos recabados de las páginas de horarios de la Facultad de Ciencias de la UNAM (Facultad). Se obtendrá un número estimado de alumnos, para cada materia y por cada hora, de las carreras del Departamento de Matemáticas. Se simularán esqueletos de horarios que se calificarán de acuerdo a ciertos criterios. Éstas simulaciones dependen de semestres anteriores, con respecto al que se quiere estimar. Se resolverá el problema de asignación de horarios por medio del algoritmo genético. Con esto se desea disminuir el tiempo que se toma actualmente el hacer tanto los esqueletos de horarios como las asignaciones de grupos en la Facultad.

1.1. Motivación

Lo que motivó la realización de este trabajo es la aportación que se puede hacer a la Facultad, la cual nos parece de gran utilidad y para el beneficio de los alumnos. Podremos obtener una disminución del tiempo que toma realizar los esqueletos y la asignación de profesores en la Facultad.

Actualmente para hacer la asignación de horarios primero se reune el comité encargado de dicha tarea a realizar manualmente los esqueletos de los horarios. Éstos se dan a conocer a los profesores y ellos eligen diferentes opciones de materias y posibles horas en las cuales les gustaría impartir sus clases. Una vez que los profesores han hecho sus solicitudes, se vuelve a hacer una o varias juntas para la asignación final de los horarios que se hace de manera manual.

Se tienen dos tipos de profesores, los de tiempo completo y los de asignatura. Los profesores de tiempo completo, por contrato, deben de cubrir ciertas horas de clase por lo que al momento de hacer la asignación se debe considerar que ellos requieren cubrir su solicitud. Finalmente se publican los horarios a los alumnos.

Una vez que los alumnos han elegido las materias que les gustaría tomar deben de ir con el profesor y él o ella les debe de firmar su tira de materias, si es que el cupo del salón lo permite. En caso de que el alumno no consiga la firma de la materia que desea, deberá buscar una segunda o tercera opción o incluso tener que meterla en algún semestre posterior.

La principal razón por la cual los profesores no firman las tiras de materias es porque el número de alumnos que desean inscribirse a su clase es mayor al número de lugares disponibles en el salón asignado. Es por ello que el trabajo que hemos realizado depende de la demanda de alumnos por materia y por horario.

1.2. Definición de conceptos

Las siguientes son las definiciones que se utilizarán a lo largo del trabajo:

Materia: Curso impartido en la Facultad por algún profesor.

Horario: Hora en la que se imparte alguna materia.

Esqueleto: Conjunto Materia-Horario.

Asignación: Conjunto Materia-Horario-Profesor.

Grupo: Clave con la que se identifica una asignación.

Turno Matutino: Comprende las clases impartidas de 7:00-14:00hrs incluyendo la clase de 14:00-15:00hrs.

Turno Vespertino: Comprende las clases impartidas de 15:00-21:00hrs incluyendo la clase de 21:00-22:00hrs.

1.3. Nomenclatura

m : Número de materias que se van a impartir, $m = 202$

p : Número de profesores que van impartir alguna materia, $p = 1223$

t : Número de horas del día, $t = 15$

i : Índice para profesores, $i \in \{1, 2, 3, \ldots, p\}$

j : Índice para materias, $j \in \{1, 2, 3, \ldots, m\}$

h : Índice para las horas del día, $h \in \{1, 2, 3 \ldots, t\}$

$U_{j,i,h}$: Utilidad de que la materia j sea impartida por el profesor i a la hora h

$x_{j,i,h}$: Variable binaria que vale 1 si la materia j es impartida por el profesor i a la hora h y cero en otro caso

$V_{j,i}$: Variable binaria que vale 1 si la materia j puede ser impartida por el profesor i y cero en otro caso

s : Semestre a simular

k : Número de semestres que se tienen como ventana de información

m_grande: Matriz en la que se guarda la información por semestres

r : Matriz *m_filtrada*, submatriz de *m_grande*

vec_sem_sig: Vector con los semestres que se van a simular *num_sim* : Número de simulaciones de la demanda de alumnos para s

E : Matriz de t renglones y m columnas. En cada entrada se tiene la información del número de alumnos simulados en los grupos al crear *mat_esqueleto*.

D : Matriz de t renglones y m columnas. En la entrada (i, j) se tiene la información de la demanda de alumnos para la hora i y la materia j.

bin_DUE : Matriz binaria de t renglones y m columnas. Tiene un 1 en la entrada (i, j) si E_{ij} o D_{ij} tienen un valor distinto de cero. Tiene un cero cuando ambas matrices (D y E) tienen un cero en la entrada (i, j).

1.4. Planteamiento del problema

En el problema de asignación de horarios se quiere asociar un profesor con una materia, un salón y un horario. Existen trabajos que han aborado este problema desde otro punto de vista, por ejemplo Yazdani, Naeri y Zeinali, en su artículo *Algorithms for university course scheduling problems* [18], proponen un modelo en el cual se toman 2 decisiones: la asignación de profesor por materia y el salón en el cual se va a impartir cada materia.

Con la función objetivo planteada en dicho modelo se desea maximizar la utilidad de que el profesor i imparta la materia j, más la utilidad de que el profesor i dé clases el día t, más la utilidad de que la materia j sea impartida en el día t. Como punto de comparación, a continuación veremos las dos diferencias principales entre su modelo y el que proponemos en este trabajo.

1) No tomamos en cuenta el día en el que se imparte la materia. Ésto porque suponemos que todas las materias se imparten de lunes a viernes, a la misma hora, en el mismo salón.

2) Deseamos maximizar la utilidad de que el profesor i imparta la materia j a la hora h.

Los elementos que consideramos en nuestro modelo son:

- Esqueletos de horario: Matriz de t renglones con las horas (7-8, 8-9, ..., 21-22) y m columnas. La entrada (i, j) contiene el número de grupos simulados de la hora i para la materia j.

- Función calificadora de esqueletos: Califica de acuerdo a qué tan bien o que tan mal se cubre la demanda de los alumnos esperados.

- Conjunto de materias: Nombres de las materias impartidas en la Facultad.

- Conjunto de profesores: Nombres de profesores de tiempo completo y de asignatura.

I) Variables de decisión:

$$x_{j,i,h} = \begin{cases} 1 & \text{si la materia } j \text{ es impartida por el profesor } i, \text{ a la hora } h \\ 0 & \text{e.o.c.} \end{cases}$$

II) Función objetivo: (se desea maximizar la utilidad)

$$\text{máx } z = \sum_{i=1}^{p}\sum_{j=1}^{m}\sum_{h=1}^{t} x_{j,i,h}U_{j,i,h} \text{ s. a}$$

III) Restricciones:

$$\begin{aligned} \sum_{i=1}^{p}\sum_{h=1}^{t} x_{j,i,h} &= 1 \quad \forall\, j &(1.1)\\ \sum_{j=1}^{m} x_{j,i,h} &\leqslant 1 \quad \forall\, i,h &(1.2)\\ \sum_{h=1}^{t} x_{j,i,h} &\leqslant V_{i,j} \quad \forall\, i,j &(1.3)\\ x_{j,i,h},\, V_{j,i} &\in \{0,1\} \quad \forall\, j,i,h &(1.4)\end{aligned}$$

Con las restricciones del tipo (1.1) aseguramos que todas las materias sean dadas. Con las del tipo (1.2) aseguramos que cada profesor no tenga más de un curso por hora. Con las del tipo (1.3) aseguramos que los profesores tengan asignadas materias que puedan impartir. Finalmente con las restricciones del tipo (1.4) se especifica que las variables utilizadas son binarias.

El conjunto de soluciones se presenta por medio de la matriz *mat_asignaciones* la cual es una matriz de tres columnas y tantos renglones como grupos se hayan simulado. En el i-ésimo renglón se tiene la información de la i-ésima materia con su respectivo profesor y horario asignados. En la Tabla 1.1 se muestra un ejemplo del resultado de la asignación.

Materia	**Profesor**	**Horario**
Inferencia Estadística	Margarita Elvira Chávez Cano	9-10
Modelos no Paramétricos y de Regresión	Jaime Vázquez Alamilla	10-11
Estadística Bayesiana	Ruth Selene Fuentes García	11-12
Modelos de Supervivencia y de Series de Tiempo	Lizbeth Naranjo Albarrán	13-14

Tabla 1.1: *Se muestra un ejemplo de la matriz mat_asignaciones que tiene 3 columnas (Materia, Profesor, Horario).*

1.5. Objetivos

El primer objetivo del trabajo es hacer dos funciones que generen:

i) Esqueletos de horarios

ii) Una asignación de profesores por materia y por horario.

Los esqueletos de horarios son utilizados para simular una posible elección de materias y horarios de los profesores. La asignación debe cubrir la demanda de alumnos estimada para el semestre siguiente. Para generar los esqueletos de horarios se simula una posible solicitud de materias y horarios de los profesores.

El segundo objetivo es disminuir el tiempo utilizado actualmente para la realización de la asignación de horarios.

1.6. Datos a analizar

Para poder realizar el análisis de los datos, hicimos 4 grupos con respecto a dos criterios. El primer criterio fue con respecto al tipo de semestre, par o impar y el segundo con respecto al turno, matutino o vespertino.

Para explicar la elección de los criterios tomamos la información de la materia *Probabilidad I*, desde el semestre 2015-1 hasta el 2020-1. Cabe aclarar que dicha materia en la carrera de Actuaría es una materia obligatoria de tercer semestre. En las siguientes subsecciones veremos el análisis de acuerdo a cada criterio.

1.6.1. Análisis por tipo de semestre: par e impar

En la Figura 1.1 vemos que la línea azul representa el número de alumnos de los semestres impares y la línea roja representa el número de alumnos de los semestres pares. Observamos que en todo momento el número de alumnos de los semestres impares es mayor al número de alumnos de los semestres pares. Ésto nos interesa porque al momento de simular debemos tomar en cuenta que el número de alumnos totales de semestres impares debe de ser siempre mayor al número total de alumnos de los semestres pares.

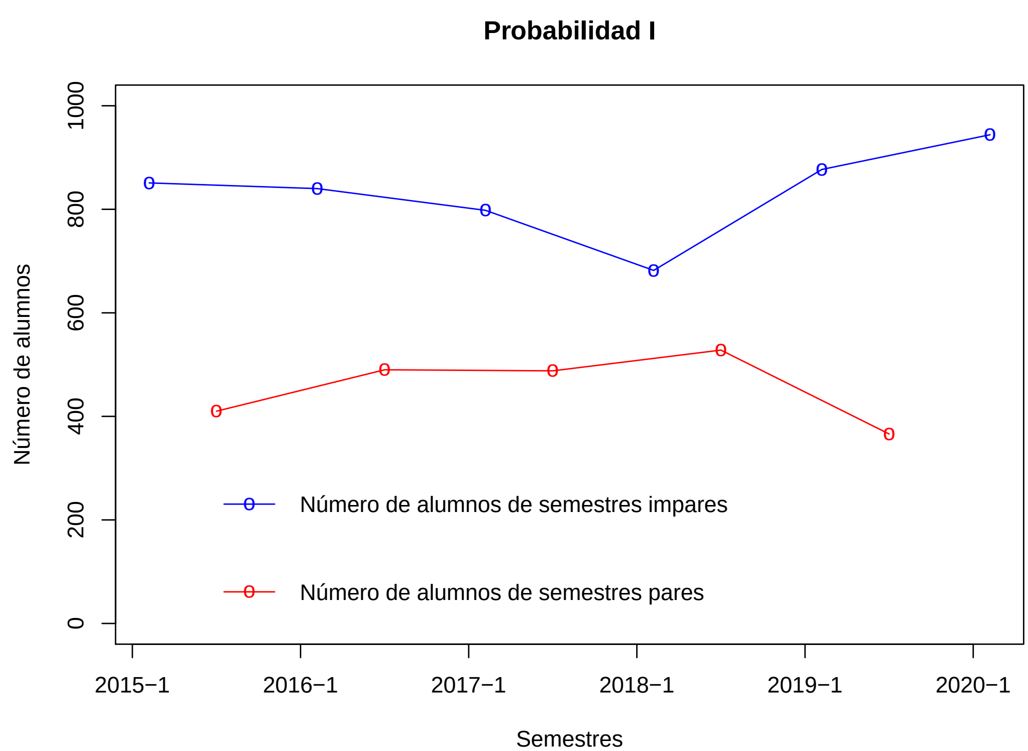


Figura 1.1: *Se muestran las series de tiempo del número de alumnos por semestres (pares e impares) de "Probabilidad I". Se puede observar que el número de alumnos de semestres impares es siempre mayor al de semestres pares.*

Continuando con los datos de *Probabilidad I*, obtuvimos la Figura 1.2. Dicha figura contiene dos histogramas, las barras rojas representan el número de alumnos por grupo de semestres pares y las barras azules representan el número de alumnos por grupo de semestres impares.

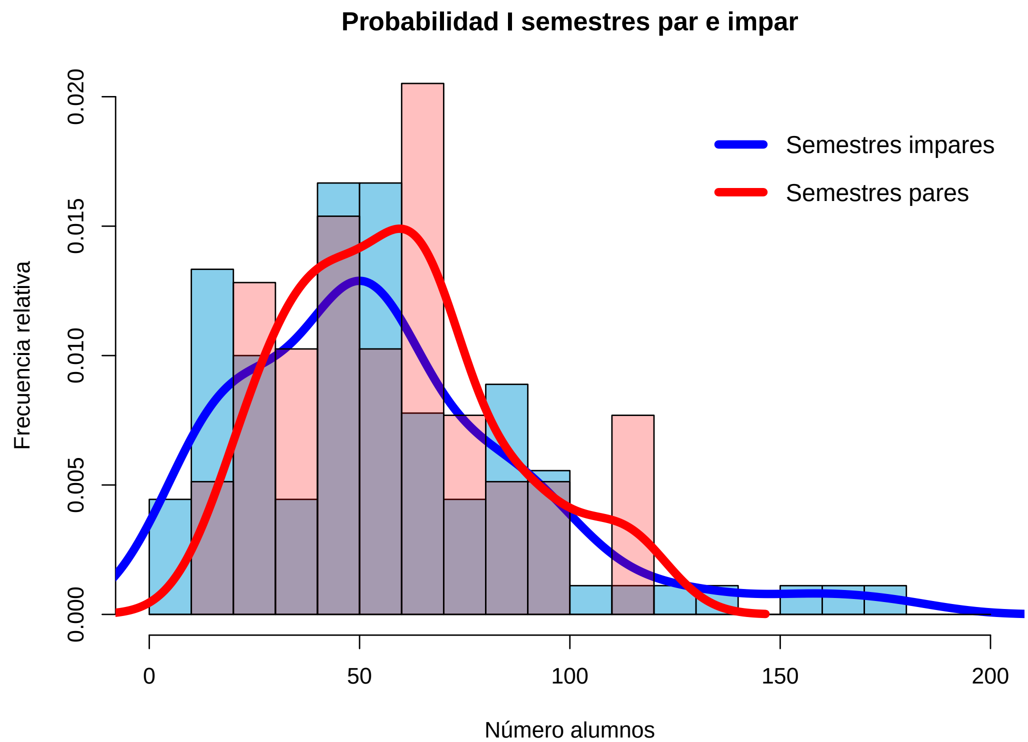


Figura 1.2: *Se muestran los histogramas del número de alumnos por semestres (pares e impares) de "Probabilidad I". Se puede observar que las densidades ajustadas son muy parecidas.*

Las líneas que se encuentran sobre los histogramas son densidades estimadas que se ajustan a los datos. Para estas aproximaciones se ajustó un kernel gaussiano con la función `density(X)` de *R*. Dicha función recibe como parámetro el vector X, con valores numéricos.

Algunos datos que se pueden obtener de las densidades vistas en la Figura 1.2 son por ejemplo que alrededor del 20 % de los grupos de los semestres pares tienen aproximadamente de 60 a 70 alumnos y que alrededor del 3 % de los grupos de los semestres impares tienen entre 150 y 180 alumnos.

1.6.2. Análisis por turno: matutino y vespertino

En la Figura 1.3 la línea azul representa el número de alumnos del turno matutino y la línea roja representa el número de alumnos del turno vespertino. Se puede observar que en todo momento el número de alumnos del turno matutino es mayor al número de alumnos del turno vespertino.

Ésto impacta en el hecho de que por semestres la varianza en el turno matutino es mucho mayor que en el turno vespertino. Lo cual indica que en el turno vespertino se tiene prácticamente el mismo número de alumnos sin importar si la materia pertenece a un semestre par o impar. Por el contrario, en el turno matutino si influye el hecho de que la materia corresponda a un semestre par o impar.

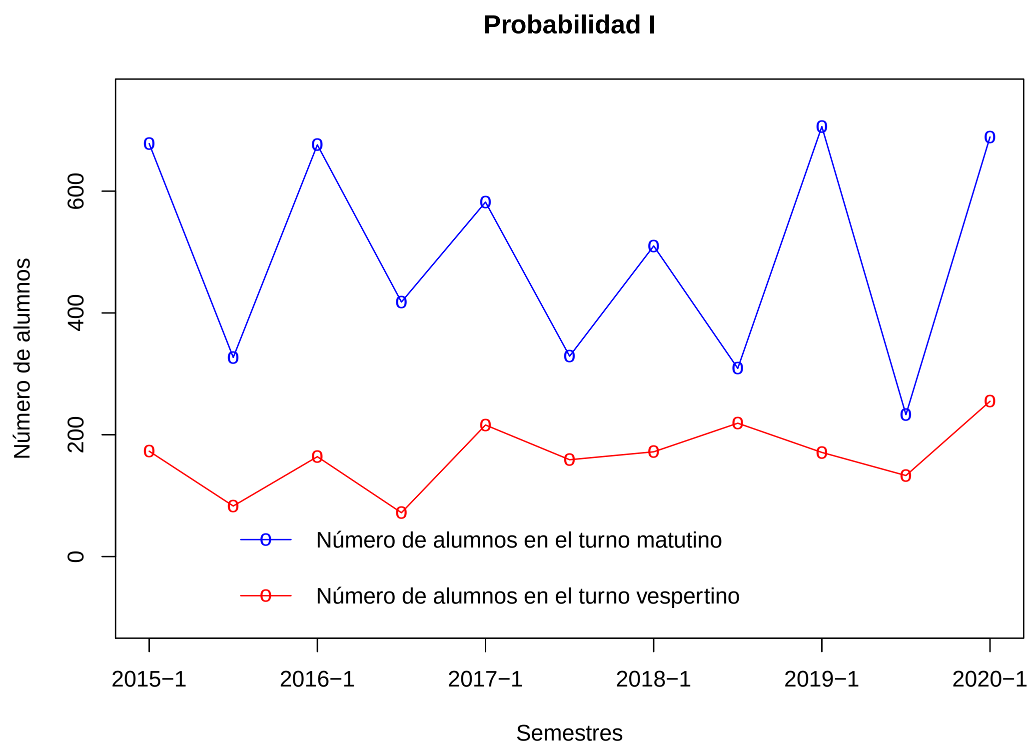


Figura 1.3: *Se muestran las series de tiempo del número de alumnos por turno (matutino y vespertino) de "Probabilidad I". Se puede ver que el número de alumnos del turno matutino es siempre mayor al número de alumnos del turno vespertino.*

En la Figura 1.4 podemos ver dos histogramas, con las densidades ajustadas correspondientes. Las barras rojas representan el número de alumnos del turno vespertino y las barras azules representan el número de alumnos del turno matutino. Notamos que en este caso las densidades ajustadas son completamente diferentes. Podemos ver que en el turno vespertino hay dos grandes concentraciones en los grupos que tienen entre 10 y 30 alumnos, así como entre 40 y 50 alumnos.

Algunos datos que se pueden obtener de las densidades ajustadas son por ejemplo que alrededor del 20 % de los grupos del turno vespertino tienen aproximadamente entre 10 y 20 alumnos y un poco más del 10 % de los grupos del turno matutino tienen entre 80 y 90 alumnos.

Con los resultados obtenidos definimos los grupos de datos G_1, G_2, G_3, G_4, para hacer los análisis estadísticos, los cuales se muestran en la Tabla 1.2.

Sem. \ Turno	**Matutino**	**Vespertino**
Impar	G_1	G_2
Par	G_3	G_4

Tabla 1.2: *Se muestran los 4 grupos obtenidos al combinar los turnos (matutino y vespertino) con los tipos de semestres (pares e impares).*

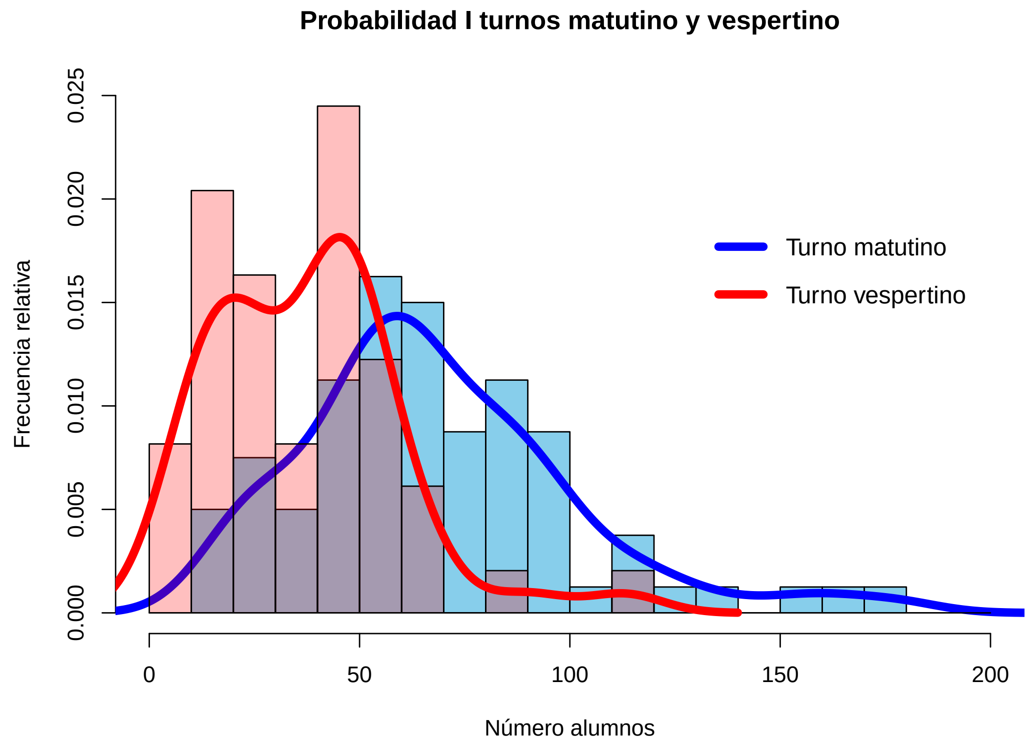


Figura 1.4: *Se muestran los histogramas del número de alumnos por turno (matutino y vespertino) de "Probabilidad I". Se puede observar que las densidades ajustadas son muy diferentes.*

Capítulo 2

Extracción de datos

La fuente de información de donde obtuvimos los datos utilizados son las páginas de los horarios de la Facultad. En la Figura 2.1 se muestra un ejemplo de dichas páginas. Cada página contiene toda la posible información de los grupos de una materia, un semestre y una carrera. Cabe mencionar que sólo tomamos en cuenta la información de las carreras del Departamento de Matemáticas, las cuales son: Actuaría, Ciencias de la Computación, Matemáticas y Matemáticas Aplicadas.

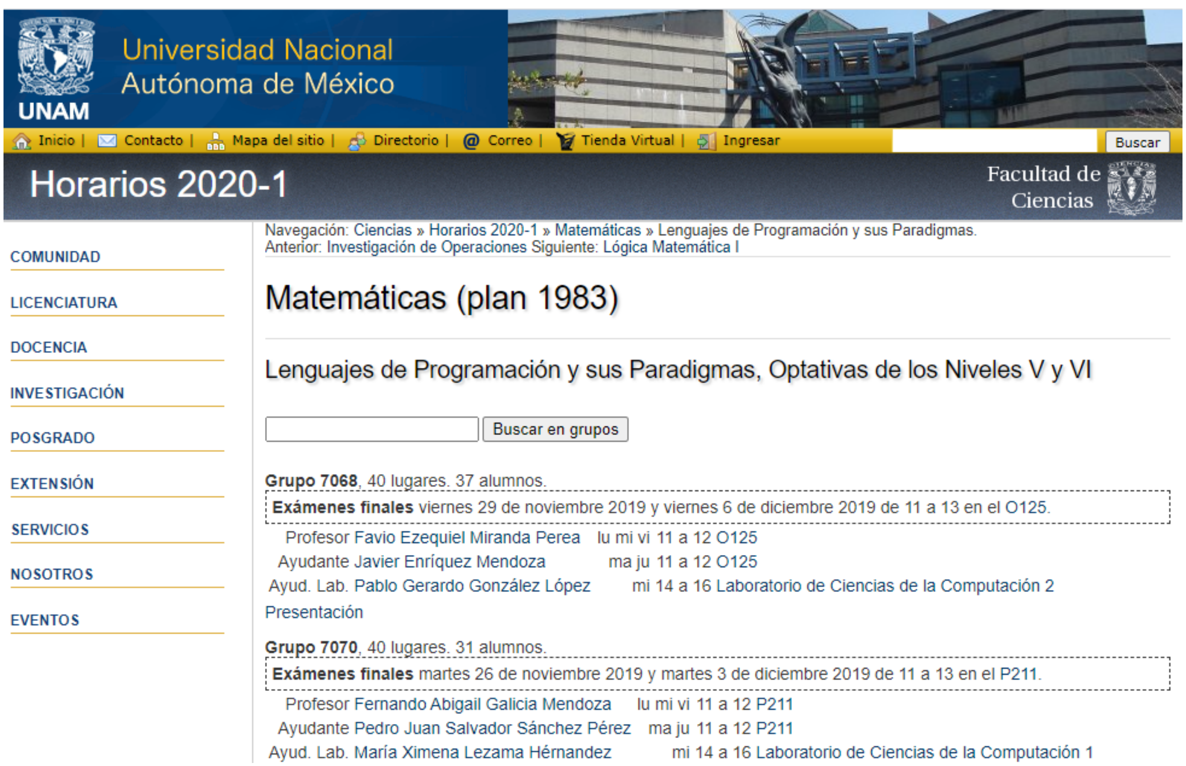


Figura 2.1: *Se muestra un ejemplo de una página de horarios de la Facultad. Se puede ver la información de los horarios de la materia "Lenguajes de Programación y sus Paradigmas", de la carrera de Matemáticas, plan 1983, del semestre 2020-1.*

La información que se puede extraer de las páginas mencionadas es: *nombre de profesores, nombre de ayudantes, salón, horario, plan, carrera, año, número de semestre, materia,* *semestre de la materia, tipo de materia* e *información de exámenes finales*.

2.1. Estructura de las URL's

Al iniciar la búsqueda de información notamos que las URL's de las páginas web de los horarios de la Facultad tienen una estructura similar. Ésto nos permitió poder realizar la búsqueda de la información de una manera automática y mucho más rápida. Observamos que la estructura que siguen las URL's mencionadas es la siguiente:

http://www.fciencias.unam.mx/docencia/horarios/a/b/c

Se tiene una raíz común para todas las páginas y al final se tienen tres números los cuales representan:

a = año y número de semestre

b = clave del plan de estudios

c = número de materia

Para este trabajo tomamos en cuenta sólo los planes de estudio vigentes hasta el semestre 2020-1. Es decir, tomamos todos los planes mostrados en la tabla Tabla 2.1, salvo el plan 1972 de Actuaría (el cual ya no está vigente). Dicha tabla muestra los planes de estudio de cada carrera con su clave correspondiente.

PLAN	CLAVE
Actuaría	
1972	214
2000	119
2006	1176
2015	2017
Ciencias de la Computación	
1994	218
2013	1556
Matemáticas	
1983	217
Matemáticas Aplicadas	
2017	2055

Tabla 2.1: *Se muestran los planes de estudio por carrera con clave. Se sustituye **b** por la clave de cada plan de estudios, en la estructura de las URL's de las páginas de la Facultad.*

Una vez identificada la estructura de las URL's pudimos realizar la búsqueda de información de manera automatizada. Originalmente decidimos que $c \in \{1, 2, 3, \ldots, 10000\}$. Después hicimos una función que genera una matriz llamada *mat_posibles_url*. La función sólo guarda las URL's que si existen.

La descripción de lo que contiene cada columna de la matriz *mat_posibles_url* la podemos ver en la Tabla 2.2. Finalmente al obtener dicha matriz, observamos que el valor máximo que toma c es 991, por lo que redujimos su conjunto de posibles valores y definimos $c \in \{1, \ldots, 1000\}$.

Col.	Nombre	Explicación	Posibles valores
1	Semestre	Semestre al que pertenece la materia (Año y semestre)	20081, ... , 20192, 20201
2	Plan	Año en el que se implementó un nuevo plan de estudios	1983, 1994, 2000, 2006, 2013, 2015, 2017
3	Materia	Clave del curso impartido	$\mathbb{N}$
4	URL	Nombres de las páginas de los horarios de la Facultad	Páginas web de la Facultad
5	Num. Grupos	Número de grupos que hay en cada página de internet	$\mathbb{N}$

Tabla 2.2: *Se muestra la descripción de las columnas de la matriz "mat_posibles_url". La matriz contiene información de cada URL existente.*

Decidimos buscar información en 25 semestres, del 2008-1 al 2020-1. Al multiplicar el número de semestres por los posibles valores de **c** obtuvimos $25 \times \mathbf{c} = 25,000$. Este valor es una aproximación del número de posibles URL's con información de los horarios de la Facultad. Notemos que no estamos contando los planes de estudio, de ser así el número supera las $170,000$ posibles URL's.

Deseamos obtener información de cada una de esas páginas. Obtener dicha información ingresando a cada una de las páginas es muy tardado por lo que es necesario hacerlo de manera automatizada. Para extraer los datos de las páginas de la Facultad utilizamos una aplicación de *Google Chrome* llamada *SelectorGadget*. En la siguiente sección, explicaremos como obtuvimos los datos de las páginas de la Facultad con dicha aplicación.

2.2. Extracción de datos con la aplicación SelectorGadget

La aplicación *SelectorGadget* permite seleccionar la información deseada y arroja una sección del código CSS de una página web. Dicho código se introduce en *R* para poder seleccionar y descargar la información deseada.

A continuación veremos los pasos que se deben de seguir para obtener el código CSS de la información seleccionada. Los colores y señalizaciones mencionados hacen referencia a la Figura 2.2.

1. Presionar el ícono de la aplicación, el cual es una lupa (señalado por la flecha roja).
2. Seleccionar la información deseada (en color verde).
3. La aplicación automáticamente selecciona todas las entradas que coindiciden (en color amarillo).
4. En caso de que se haya seleccionado más información de la deseada entonces dar click sobre la información excedente (en color rojo).
5. En el cuadro de texto, la aplicación arroja la sección del código CSS correspondiente a la información seleccionada. También muestra el número de entradas selecionadas (en óvalos rojos).

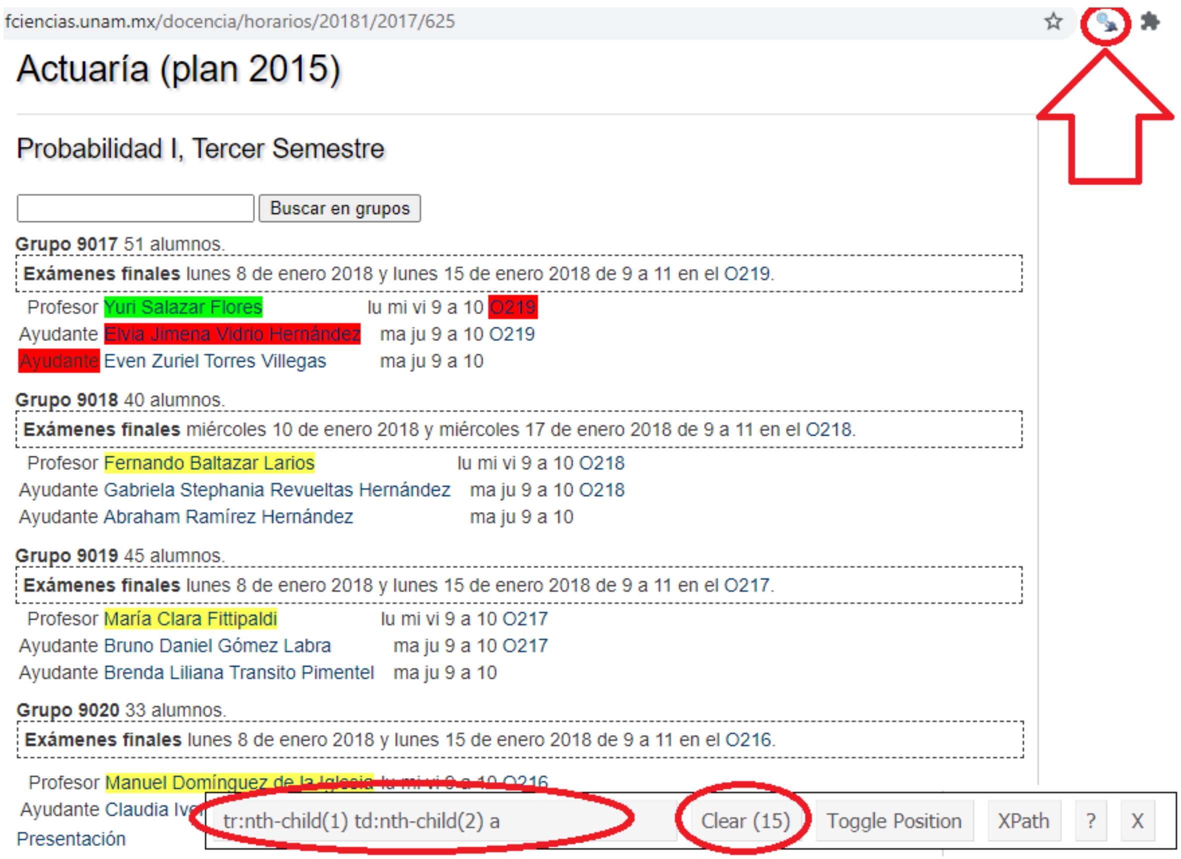


Figura 2.2: *En esta figura se muestra cómo se ve una página de horarios de la Facultad al usar la aplicación SelectorGadget mientras se selecciona la información que deseamos extraer.*

En el ejemplo mostrado en la Figura 2.2, se seleccionaron 15 entradas correspondientes a los nombres de los profesores en una página con la información de la materia *Probabilidad I*, en el plan 2015 de Actuaría. Los pasos a seguir son los mismos sin importar la información que se desea obtener, lo único que cambia es el código CSS que arroja la aplicación.

Sabiendo ésto, hicimos un ciclo en *R* que recorre todas las posibles combinaciones de las URL's. Encontramos 3 principales tipos de grupos, los cuales describiremos en la siguiente sección.

2.3. Tipos de grupos en las páginas web de la Facultad de Ciencias

Cada uno de los 3 tipos de grupos encontrados contienen información similar. Hicimos la separación de acuerdo a sus diferencias. Cabe mencionar que en este trabajo consideramos como semestre actual al semestre $2020-1$. En todos los grupos se puede encontrar la información del nombre de profesor, nombre del o de los ayudantes, salón, horario y el número de alumnos inscritos en el grupo.

a) En el tipo **A** se tienen las páginas correspondientes al semestre actual. Este tipo de grupo tiene la información del número de lugares diponibles por salón, pero no contiene la información de los exámenes finales, porque se considera que el semestres aún está en curso y aún no termina. En la Figura 2.3 podemos ver un ejemplo de este tipo de grupo.

Grupo 9301, 129 lugares. 84 alumnos.

Profesor José Luis Navarro Urrutia lu mi vi 13 a 14 Aula Magna I

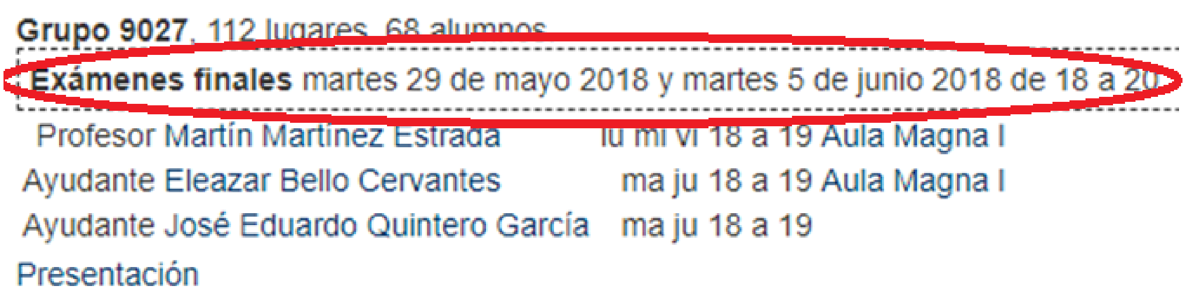
Ayudante Luz Candy Becerril Palacios ma ju 13 a 14 Aula Magna I

Ayudante Gabriela Yaneth Romo Cordoba ma ju 13 a 14

Ayudante Adrián Gallardo Pacheco ma ju 13 a 14

Figura 2.3: *Se muestra un ejemplo de un grupo de tipo A. Corresponde al semestre actual (2020-1).*

b) En el tipo **B** se tienen las páginas correspondientes a semestres entre el $2018-2$ y el $2019-2$ (semestre anterior al actual). En este tipo de grupos se tiene información del número de lugares disponibles por salón y la información de los exámenes finales. Ésto porque son semestres que ya finalizaron. En la figura Figura 2.4 encontramos un ejemplo de este tipo de grupo.



Grupo 9027, 112 lugares. 68 alumnos.

Exámenes finales martes 29 de mayo 2018 y martes 5 de junio 2018 de 18 a 20.

Profesor Martín Martínez Estrada lu mi vi 18 a 19 Aula Magna I

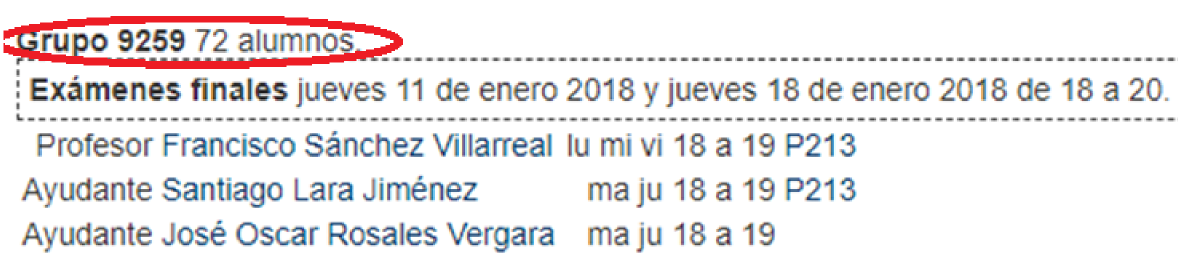
Ayudante Eleazar Bello Cervantes ma ju 18 a 19 Aula Magna I

Ayudante José Eduardo Quintero García ma ju 18 a 19

Presentación

Figura 2.4: *Se muestra un ejemplo de un grupo de tipo B. Corresponde a semestres ya finalizados, desde el 2018-2 hasta el 2019-2.*

c) En el tipo **C** se tienen las páginas correspondientes a semestres anteriores al $2018-1$, incluyéndolo. Este tipo de grupos tiene información de los exámenes finales, pero no contiene la información del número de lugares disponibles por salón. En la figura Figura 2.5 podemos ver un ejemplo.



Grupo 9259 72 alumnos.

Exámenes finales jueves 11 de enero 2018 y jueves 18 de enero 2018 de 18 a 20.

Profesor Francisco Sánchez Villarreal lu mi vi 18 a 19 P213

Ayudante Santiago Lara Jiménez ma ju 18 a 19 P213

Ayudante José Oscar Rosales Vergara ma ju 18 a 19

Figura 2.5: *Se muestra un ejemplo de un grupo de tipo C. Corresponde a semestres ya finalizados, anteriores al semestre 2018-1, incluyéndolo.*

2.4. Limpieza de base de datos

Se puede encontrar que, en general, cuando uno realiza la limpieza de datos se hace el 80 % del análisis de los datos. Es en ese momento en donde se encuentran los diferentes problemas que se pueden presentar. Se pueden encontrar posibles errores en los datos, información incompleta o valores poco comunes de acuerdo al comportamiento observado. Los problemas que encontramos al limpiar los datos se desglosan en las siguientes subsecciones.

2.4.1. Problemas de falta de información

Encontramos diferentes páginas que tenían grupos sin información e incluso páginas sin información alguna. Para guardar la información consideramos sólo los grupos que al menos tenían: nombre de profesor, número de alumnos inscritos y horario. A continuación se muestran varios ejemplos con los diferentes casos de falta de información encontrados.

- En la Figura 2.6 vemos un ejemplo de páginas en las cuales se tiene el nombre de la materia, pero no hay información de algún grupo: `http://www.fciencias.unam.mx/docencia/horarios/20081/1556/803`

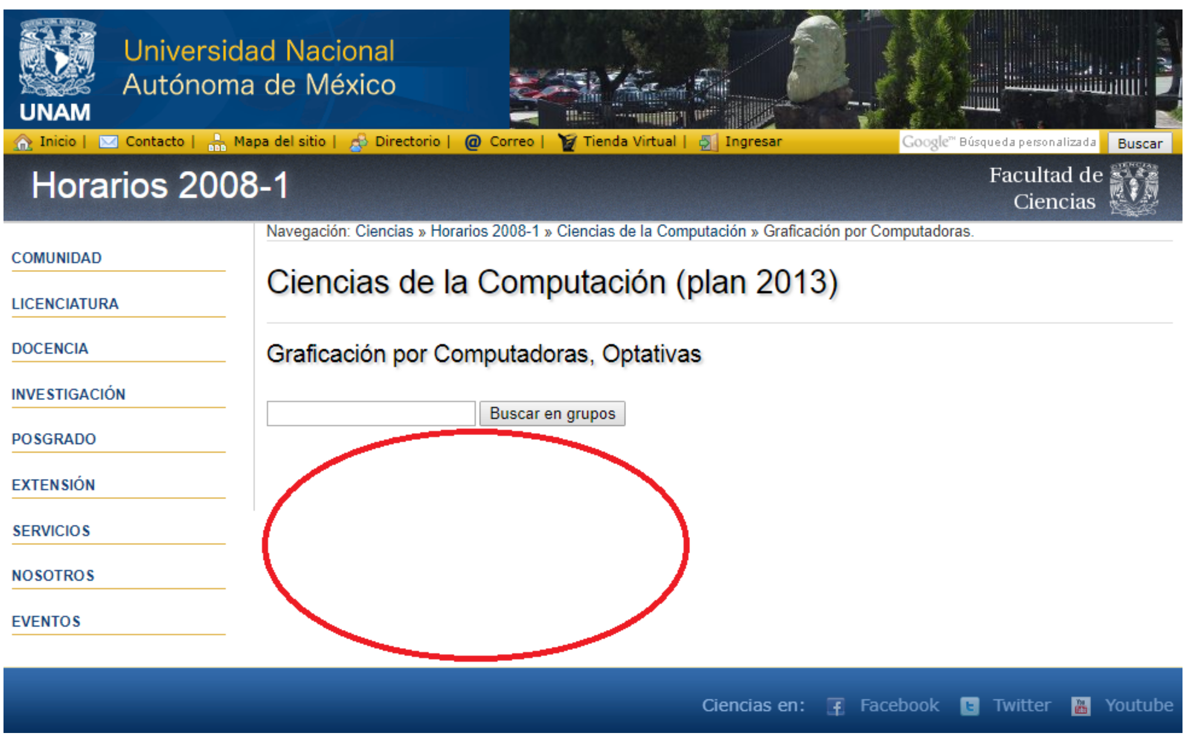


Figura 2.6: *Se muestra un ejemplo de página web en blanco para la materia "Graficación por Computadoras". En este tipo de páginas no encontramos información de ningún grupo.*

- En la Figura 2.7 encontramos un ejemplo de páginas que no tienen información del salón: `http://www.fciencias.unam.mx/docencia/horarios/20081/119/4`

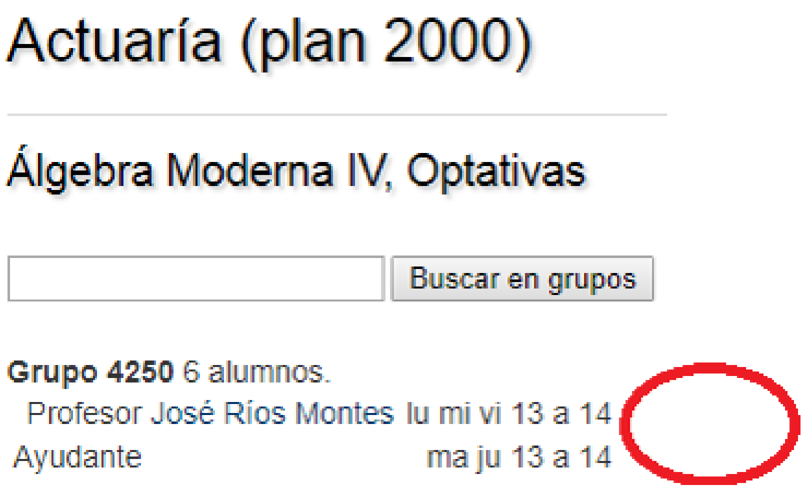


Figura 2.7: *Se muestra un ejemplo de grupo sin información de salón. En este tipo páginas no se muestra el salón en el que se imparte la clase.*

- En la Figura 2.8 tenemos un ejemplo de páginas que tienen grupos sin información del número de alumnos inscritos: `http://www.fciencias.unam.mx/docencia/horarios/20112/119/630`

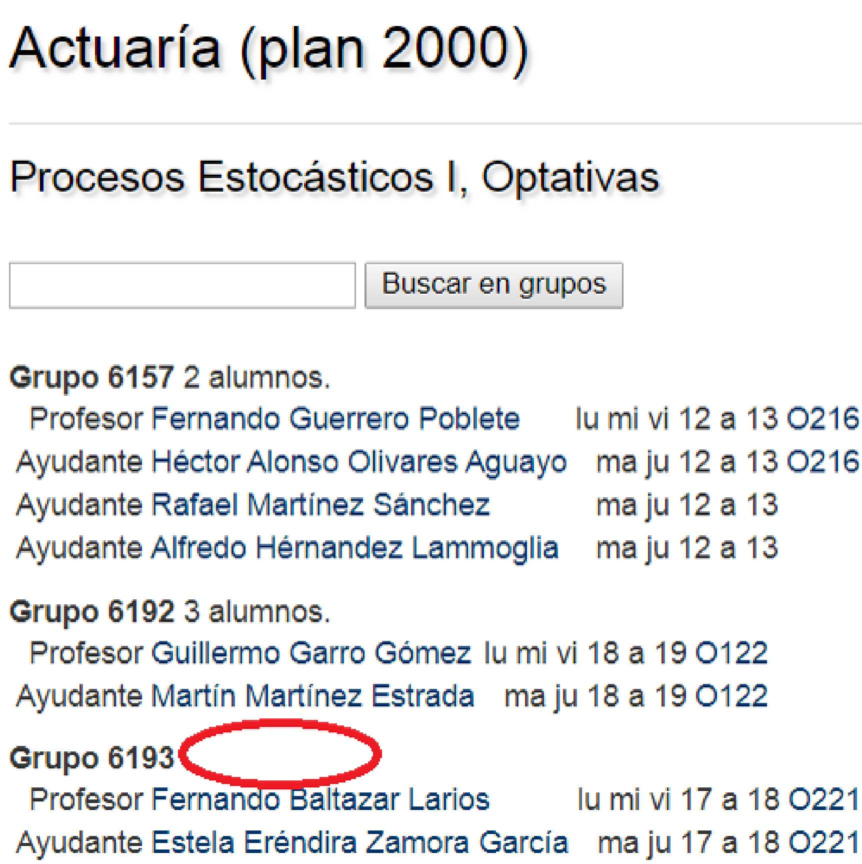


Figura 2.8: *Se muestra un ejemplo de grupo sin información de alumnos. En este tipo páginas encontramos grupos que no tienen el número de alumnos inscritos.*

- En la Figura 2.9 vemos un ejemplo de páginas que tienen grupos sólo con el horario.

No tienen nombre del profesor, salón, ayudante, número de alumnos, ni lugares disponibles: `http://www.fciencias.unam.mx/docencia/horarios/20091/119/841`

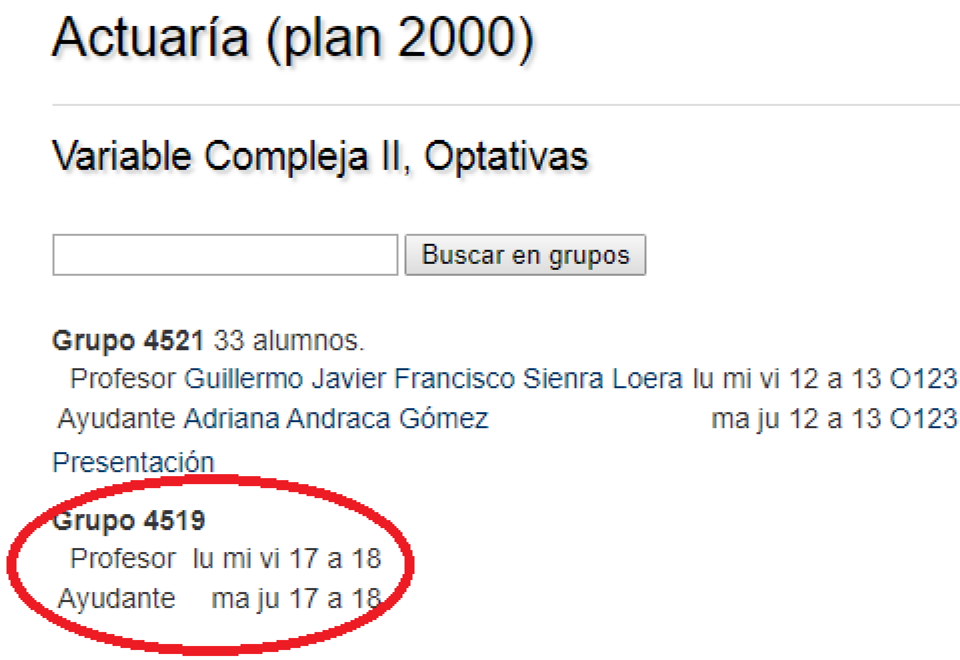


Figura 2.9: *Se muestra un ejemplo de grupo sólo con horario. En este tipo páginas existen grupos que no tienen información del profesor o salón ni del número de alumnos inscritos. Sólo tienen la clave del grupo y el horario.*

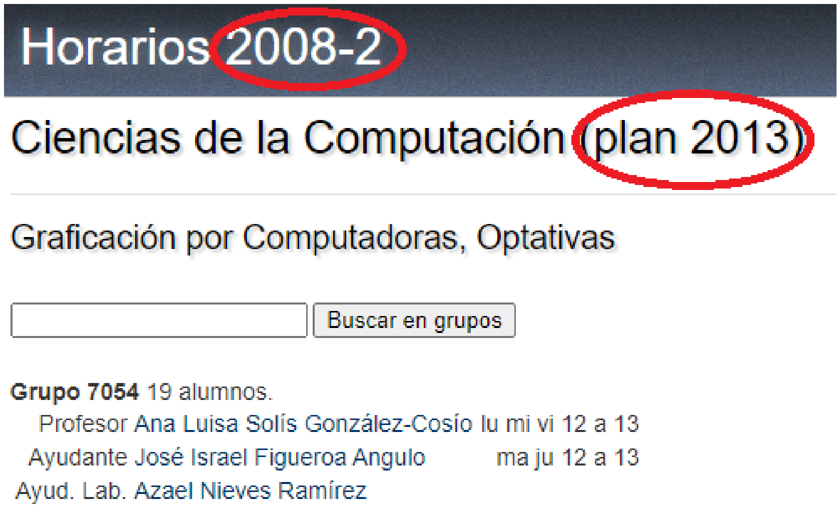
2.4.2. Problemas de información repetida

Dentro de los problemas de información repetida, para guardar la información, juntamos aquellas clases que provenían del mismo grupo. A continuación presentamos los casos que encontramos con el problema de tener información repetida.

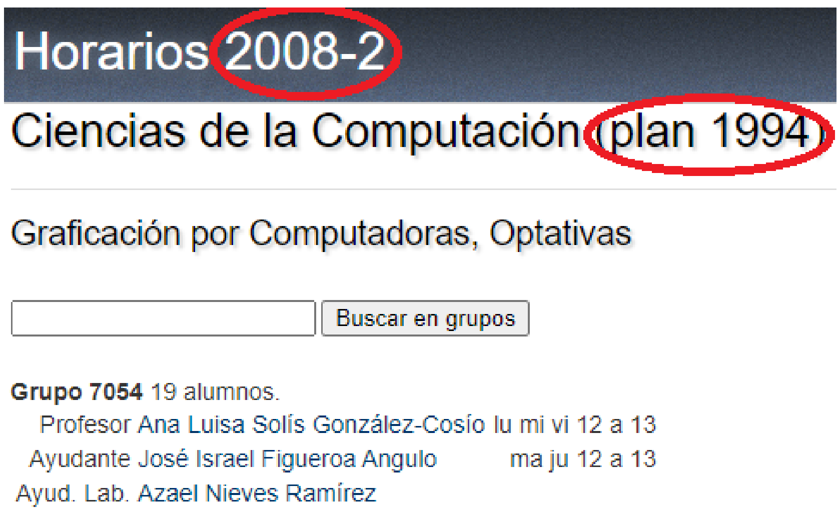
- El número del plan de estudios corresponde al año en que entró en vigencia el plan. Por ejemplo, si se tiene un plan 2015 en Actuaría, entonces dicho plan comenzó a tener vigencia a partir del año 2015. Debido a ésto no debería de existir un horario con un plan posterior al año del semestre.

  En la subfigura (a) de la Figura 2.10 podemos ver una materia de la carrera de Ciencias de la Computación del semestre 2008-2, con el plan 2013, lo cual no es cronológicamente correcto: `http://www.fciencias.unam.mx/docencia/horarios/20082/1556/803`.

  En la subfigura (b) de la misma figura, vemos la información de la misma materia y del mismo grupo pero con el plan 1994: `http://www.fciencias.unam.mx/docencia/horarios/20082/218/803`.



(a) *Plan de estudios posterior*



(b) *Plan de estudios correspondiente*

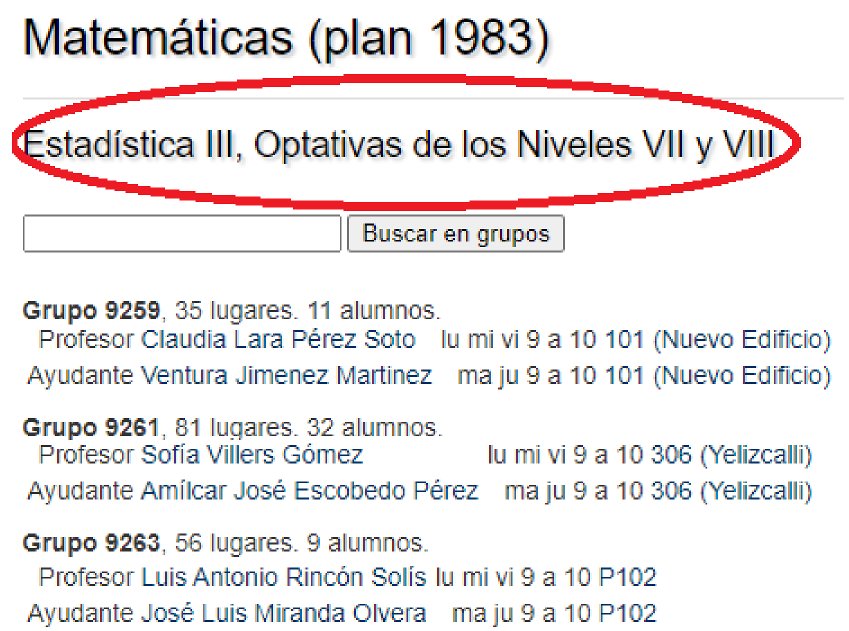
Figura 2.10: *Se muestra un ejemplo de información repetida por los planes de estudio. No deberían de existir grupos con planes posteriores al año del semestre en el que se busca información.*

- Encontramos grupos correspondientes a una misma materia con nombres distintos para diferentes carreras.

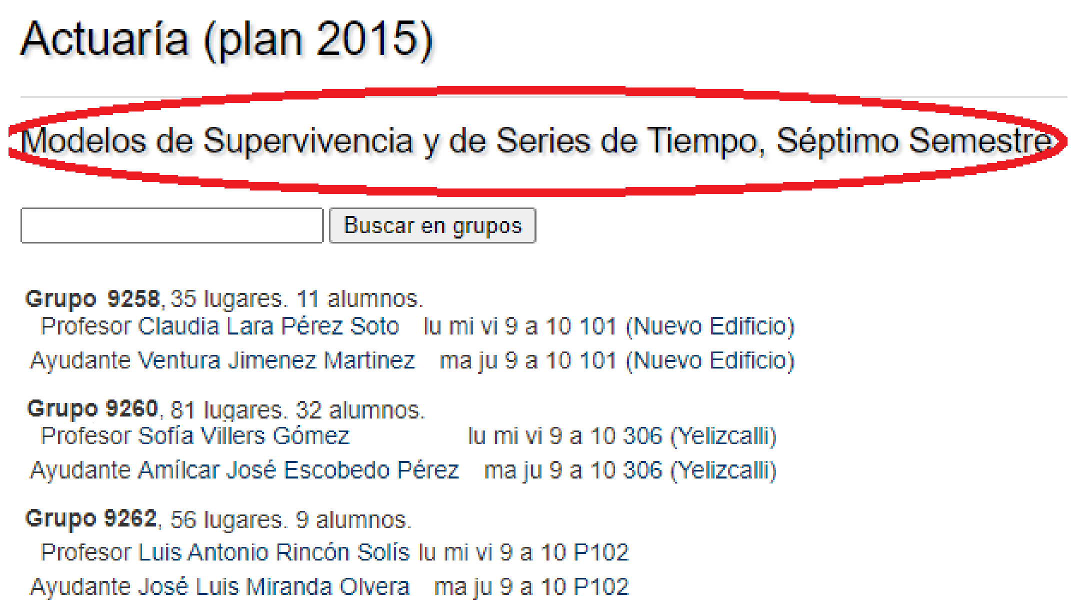
  En la subfigura (a) de la Figura 2.11 vemos un ejemplo con la información de la materia *Estadística III*, para la carrera de Matemáticas plan 1983: `http://www.fciencias.unam.mx/docencia/horarios/20201/217/1712`.

  En la subfigura (b) de la figura mencionada se muestra la información de la materia *Modelos de Superviciencia y de Series de Tiempo*, para la carrera de Actuaría plan 2015: `http://www.fciencias.unam.mx/docencia/horarios/20201/2017/1739`.

  Notamos que la información en ambos ejemplos es la misma. Sólo cambian las claves de los grupos y el nombre de las materias.



(a) *Matemáticas plan 1983*



(b) *Actuaría plan 2015*

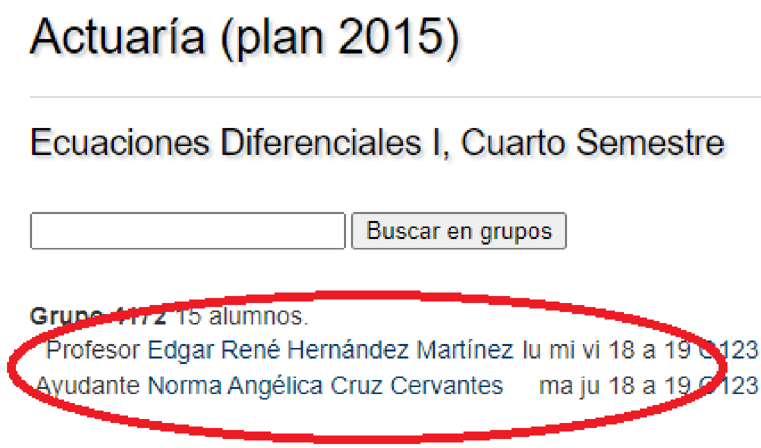
Figura 2.11: *Se muestra un ejemplo de información repetida por materia con nombres distintos. En estos casos se tienen materias que tienen nombres diferentes de acuerdo a la carrera o plan de estudios.*

- Encontramos grupos con profesores que imparten dos o más clases distintas en el mismo horario y diferente salón.

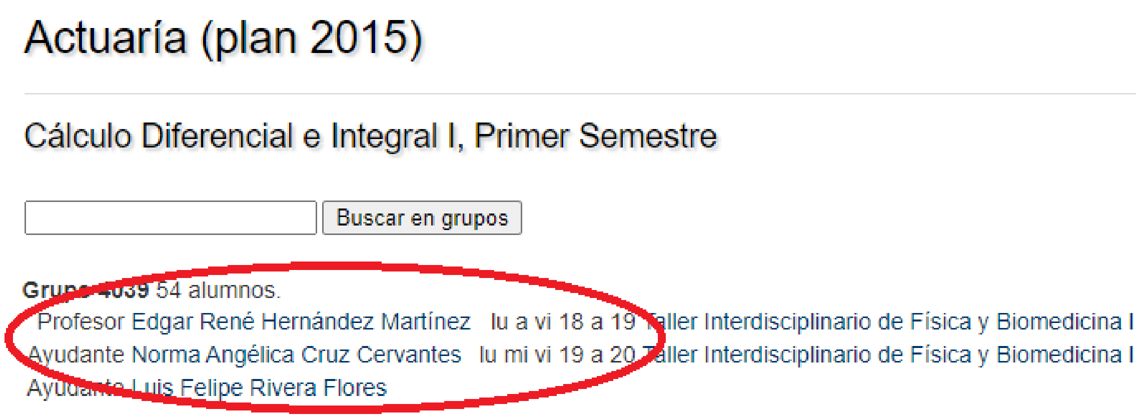
  En la subfigura (a) de la Figura 2.12 se observa un ejemplo con la información de la materia *Ecuaciones Diferenciales I*: `http://www.fciencias.unam.mx/docencia/horarios/20111/2017/162`.

  En la subfigura (b) de la figura mencionada se muestra la información de la materia *Cálculo Diferencial e Integral I*: `http://www.fciencias.unam.mx/docencia/horarios/20111/2017/91`.

Las materias mencionadas son diferentes, pero las clases comienzan a la misma hora, *Ecuaciones Diferenciales I* de 18-19hrs y *Cálculo Diferencial e Integral I* de 18-20hrs. Dado que se tiene la misma ayudante pudiera ser que se intercambien las horas, pero no se puede asignar más de una clase a la misma hora al mismo profesor.



(a) *Ecuaciones Diferenciales I*



(b) *Cálculo Diferencial e Integral I*

Figura 2.12: *Se muestra un ejemplo de información repetida por tener un mismo profesor impartiendo materias distintas, en distinto salón, a la misma hora.*

2.4.3. Otros problemas al extraer información

En algunos de los problemas que surgieron, encontramos detalles particulares que tuvimos que resolver caso por caso. Ésto para poder guardar la información de manera adecuada. A continuación se presentan los diferentes casos encontrados:

- Dentro de la obtención de datos del número de alumnos, no se lee la información cuando se tiene *“Un alumno”*, ya que no se reconoce el texto *“Un”* como el número 1. En la Figura 2.13 vemos un ejemplo de este caso.

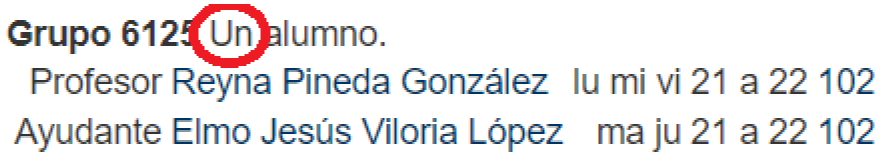


Figura 2.13: *Se muestra un ejemplo de grupo con el texto "Un" y no el número* 1.

Para resolver este problema se identificó la variable tipo *string* igual a *"Un"* para convertir la información y así poder utilizar los datos obtenidos.

- El algoritmo supone que todas las clases duran una hora y no se consideran las medias horas: `http://www.fciencias.unam.mx/docencia/horarios/20172/1556/820`. En la Figura 2.14 mostramos un ejemplo en donde se considera que esa materia inicia a las 18hrs.

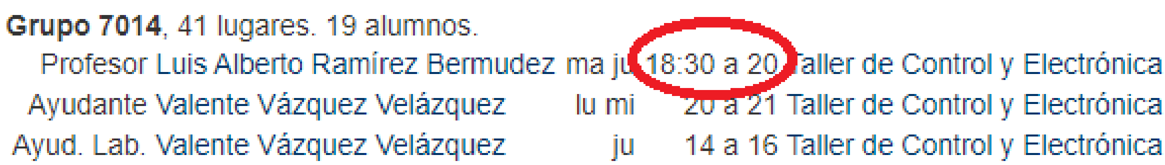


Figura 2.14: *Se muestra un ejemplo de grupo con medias horas. Se considera que las materias inician en horas enteras y no a las medias horas.*

- Se tienen materias con múltiples horarios: `http://www.fciencias.unam.mx/docencia/horarios/20181/2055/1323`. En estos casos sólo se registran los horarios y salones en los que los profesores imparten su clase, no se toman en cuenta las clases impartidas por los ayudantes.

  En la Figura 2.15 tenemos un ejemplo de este caso en donde el profesor imparte su clase los lunes, miércoles y viernes de 13-14hrs en el salón O215, hay una ayudantía los martes y jueves de 13-14hrs en el salón O215 y otra ayudantía los martes de 11-13hrs en el salón 304 (Yelizcalli). Se considera que esta materia inicia a las 13hrs y se imparte en el salón O215.

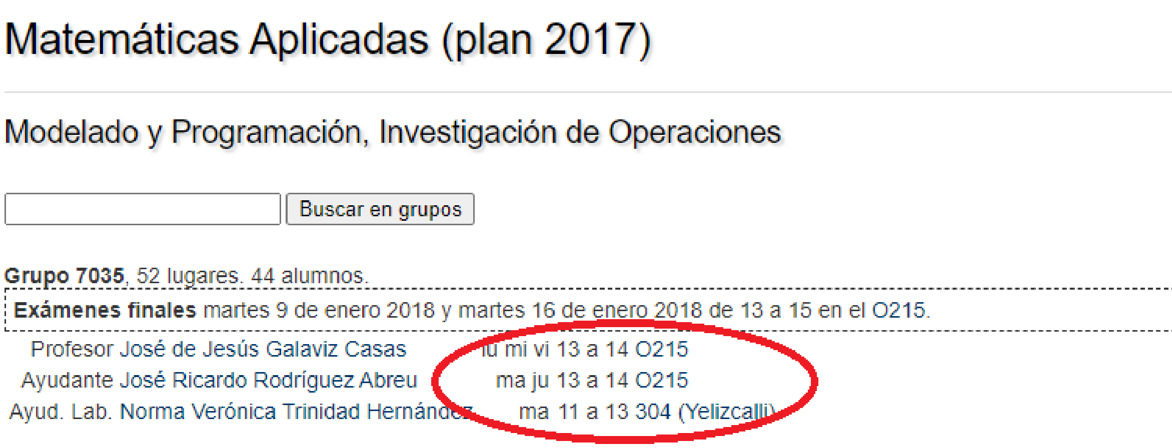


Figura 2.15: *Se muestra un ejemplo de grupo con horarios múltiples. En estos grupos sólo se toman en cuenta los horarios y salones en los que los profesores imparten clase.*

- Las materias de inglés no se imparten todos los días de la semana, en algunos casos se imparten clases en línea: `http://www.fciencias.unam.mx/docencia/horarios/20202/2017/1135`. Se registran únicamente los horarios de los días en que se imparten las clases presenciales. En la Figura 2.16 mostramos un ejemplo de este caso.

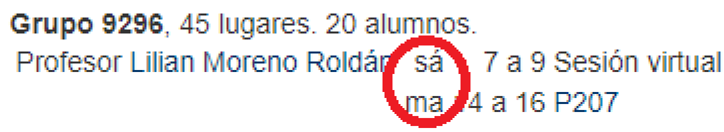


Figura 2.16: *Se muestra un ejemplo de un grupo de inglés. Las clases no se imparten todos los días. Hay sesiones virtuales. Sólo se toma en cuenta el horario de las clases presenciales.*

- Se tienen grupos que no tienen la misma estructura que los tipos de grupos **A**, **B** y **C** definidos en la Sección 2.3: `http://www.fciencias.unam.mx/docencia/horarios/20201/2017/872`, debido a ello el código CSS utilizado no sirve para obtener toda la información que se puede obtener del grupo. En la Figura 2.17 tenemos un ejemplo de este caso en donde no se lee adecuadamente el número de alumnos inscritos en el grupo.

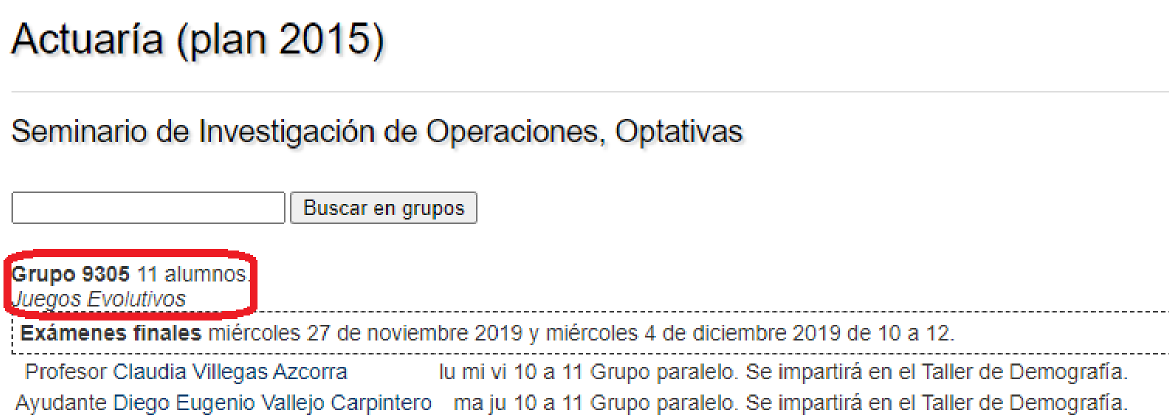


Figura 2.17: *Se muestra un ejemplo de grupo con estructura diferente. En estos casos no se extrae adecuadamente la información de los grupos porque el código CSS utilizado no corresponde a este tipo de grupos.*

2.5. Matrices de datos

Una vez que se realizó el proceso de limpieza de los datos obtenidos, éstos se guardaron, por semestre, en matrices llamadas *m_grande*. Los nombres de sus columnas con su respectiva explicación y posibles valores, se muestran en la siguiente tabla. Cabe aclarar que la abreviatura *CdC* denota *Ciencias de la Computación* y *MatAp* denota *Matemáticas Aplicadas*.

Col.	**Nombre**	**Explicación**	**Posibles valores**
1	Materia	Nombre de algún curso impartido en la Facultad	“Probabilidad I”

La tabla continúa en la siguiente página

Col.	**Nombre**	**Explicación**	**Posibles valores**
2	Profesor	Nombre de la persona que va a impartir alguna materia	“Arrigo Coen Coria”
3	Horario	Hora en la que se imparte alguna materia	“7 a 8”, ... , “21 a 22”
4	horario_num	Valores de la columna Horario en variables tipo *numeric*	$7, 8, 9, \ldots, 20, 21$
5	Lugares	Espacios disponibles por salón	$\mathbb{N}$
6	Alumnos	Número de estudiantes inscritos por grupo	$\mathbb{N}$
7	Salón	Espacio físico en el que se imparte alguna materia	“O218”, ..., “P105”
8	Grupo	Clave con la que se identifica una asignación	4489, 6114, ...
9	Carrera	Nombre de alguna carrera de la Facultad	“Actuaría”, “Matemáticas”,...
10	Plan	Año en el que se implementó un nuevo plan de estudios	1983, 1994, ..., 2017
11	Semestre	Semestre al que pertenece la materia (Año y semestre par o impar)	20081, ... , 20192, 20201
12	Cambios	Clave que indica los cambios que se le han hecho al grupo	$\mathbb{N}$
13	Turno	Matutino: 7:00-14:00hrs, Vespertino: 15:00-21:00	M,V
14	Semestre_de_materia	Semestre en el que el plan de estudios dicta que se lleva esa materia	“Primer Semestre”, ... , “Optativas”
15	url	Nombre de la página de los horarios de la Facultad correspondiente al grupo	url’s de la Facultad
16	Act2000	Columna binaria, indica si el grupo pertenece a la carrera de Actuaría, plan 2000	$0, 1$
17	Act2006	Columna binaria, indica si el grupo pertenece a la carrera de Actuaría, plan 2006	$0, 1$
18	Act2015	Columna binaria, indica si el grupo pertenece a la carrera de Actuaría, plan 2015	$0, 1$
19	CdC1994	Columna binaria, indica si el grupo pertenece a la carrera de CdC, plan 1994	$0, 1$
20	CdC2013	Columna binaria, indica si el grupo pertenece a la carrera de CdC, plan 2013	$0, 1$
21	Mat1983	Columna binaria, indica si el grupo pertenece a la carrera de Matemáticas, plan 1983	$0, 1$

La tabla continúa en la siguiente página

Col.	**Nombre**	**Explicación**	**Posibles valores**
22	MatAp2017	Columna binaria, indica si el grupo pertenece a la carrera de MatAp, plan 2017	$0, 1$
23	NomMat_Act2000	Indica el nombre de las materia correspondiente a la carrera de Actuaría plan 2000	Nombres de materias de la Facultad
24	NomMat_Act2006	Indica el nombre de las materia correspondiente a la carrera de Actuaría plan 2006	Nombres de materias de la Facultad
25	NomMat_Act2015	Indica el nombre de las materia correspondiente a la carrera de Actuaría plan 2015	Nombres de materias de la Facultad
26	NomMat_CdC1994	Indica el nombre de las materia correspondiente a la carrera de CdC plan 1994	Nombres de materias de la Facultad
27	NomMat_CdC2013	Indica el nombre de las materia correspondiente a la carrera de CdC plan 2013	Nombres de materias de la Facultad
28	NomMat_Mat1983	Indica el nombre de las materia correspondiente a la carrera de Matemáticas plan 1983	Nombres de materias de la Facultad
29	NomMat_MAp2017	Indica el nombre de las materia correspondiente a la carrera de MatAp plan 2017	Nombres de materias de la Facultad
30	URL_Act2000	Indica la URL correspondiente a la carrera de Actuaría plan 2000	url de la Facultad
31	URL_Act2006	Indica la URL correspondiente a la carrera de Actuaría plan 2006	url de la Facultad
32	URL_Act2015	Indica la URL correspondiente a la carrera de Actuaría plan 2015	url de la Facultad
33	URL_CdC1994	Indica la URL correspondiente a la carrera de CdC plan 1994	url de la Facultad
34	URL_CdC2013	Indica la URL correspondiente a la carrera de CdC plan 2013	url de la Facultad
35	URL_Mat1983	Indica la URL correspondiente a la carrera de Matemáticas plan 1983	url de la Facultad
36	URL_MAp2017	Indica la URL correspondiente a la carrera de MatAp plan 2017	url de la Facultad
37	Num_materia	Número de materia de acuerdo al vector *vec_nom_materias*	$\mathbb{N}$

Tabla 2.3: *En esta tabla se describe el contenido de las columnas de las matrices "m_grande". En ellas se guarda la información de las páginas de la Facultad por semestres.*

La columna *Cambios*, va a guardar todos los cambios que ha tenido cada grupo. El significado de los números que pueden aparecer en esa columna se explican a continuación:

(1) Grupos que se revisaron individualmente por tener detalles particulares.

(2) Se anotaron los días en los que se imparte la materia, en la columna *Horario*, por ejemplo cuando había conflicto debido a que el profesor impartía más de una materia a la misma hora, al revisar el caso se encontró que los días en los que se impartía la clase era distinto.

(3) Se eliminaron los grupos repetidos, al juntar la información en un mismo grupo.

(4) Páginas que no tienen información del salón.

(5) Actualización del número de materia por cambio de nombre o agrupamiento de materias.

Capítulo 3

Análisis estadístico

Debido a la naturaleza de los datos, las herramientas elegidas para realizar un análisis estadístico de los datos fueron las series de tiempo. A continuación se describe su definición y aplicación para explicar el motivo de la elección de dichas herramientas estadísticas.

Definimos a una serie de tiempo como una secuencia de observaciones X_t ordenadas cronológicamente. Los datos al tiempo presente dependen de las observaciones anteriores, es decir existe una depencia de X_t con $\{X_{t-1}, X_{t-2}, X_{t-3}, \ldots, X_2, X_1, X_0, \ldots\}$.

Denotamos a una serie de tiempo como:

$$X_t = m_t + s_t + y_t, \tag{3.1}$$

donde las componentes de la serie de tiempo (m_t, s_t, y_t) tienen las siguientes propiedades:

- Tendencia (m_t): Se le llama tendencia al cambio, a largo plazo, del promedio de los datos. El cambio puede ser creciente o decreciente.
- Estacionalidad (s_t): Se llama variación estacional a las fluctuaciones periódicas que tiene una serie de tiempo. La longitud de cada periodo es constante y por lo general menor o igual a un año, por ejemplo semanal, mensual o semestral.
- Aleatoriedad (y_t): También llamada componente irregular, son series de residuales que pueden o no ser aleatorios.

Chatfield y Xing, en su libro *The Analysis of Time Series An Introduction with R* [3], nos indican que existen 2 tipos de variación estacional:

- Aditiva: Se dice que la estacionalidad es aditiva cuando la longitud de cada periodo es constante año con año.
- Multiplicativa: Se dice que la estacionalidad es multiplicativa cuando la longitud de cada periodo es directamente proporcional a la media de los datos de la serie de tiempo.

Con estos tipos de variaciones se forman 3 modelos de estacionalidad:

1. Aditivo: En este modelo se tiene variación estacional aditiva. Se utiliza cuando la varianza o la desviación estándar de la serie de tiempo se mantienen constantes a lo largo

del tiempo. El modelo aditivo se denota como:

$$X_t = m_t + s_t + y_t. \tag{3.2}$$

2. Multiplicativo: En este modelo se tiene variación estacional multiplicativa. Se utiliza cuando la varianza o la desviación estándar de los datos cambian a través del tiempo. Su variabilidad puede ser mayor o menor conforme pasa el tiempo. El modelo multiplicativo se denota como:

$$X_t = m_t s_t y_t. \tag{3.3}$$

3. Mixto: Este modelo se utiliza cuando se tiene variación estacional multiplicativa pero la variabilidad de la componente irregular se mantiene constante a lo largo del tiempo. El modelo mixto se denota como:

$$X_t = m_t s_t + y_t. \tag{3.4}$$

Los objetivos principales al hacer el análisis de una serie de tiempo son:

- Describir: Leer datos en una tabla es mucho más tardado y en algunas ocasiones más complicado que observar una gráfica de los datos que se tienen. Las gráficas ayudan a ver de una manera más inmediata el comportamiento que tienen los datos y es posible observar si la serie de tiempo tiene alguna tendencia o estacionalidad. También se puede ver la posible falta de información o valores atípicos.

- Predecir: Teniendo una serie de tiempo se desea conocer qué va a pasar en el futuro. Es conveniente tener varios periodos de información para que la predicción sea lo más acertada posible.

Las áreas en las que se pueden aplicar las series de tiempo son por ejemplo en economía, demografía, finanzas, medio ambiente o medicina. En estas áreas, algunos ejemplos de su aplicación son: precios de acciones diarios, niveles de producción en la agricultura mensuales, medición del sonido por segundos, barriles de petróleo producidos al año, electrocardiogramas, medición de terremotos, tasa de mortalidad, tasa de natalidad, entre otros.

3.1. Análisis estadístico básico

En esta sección haremos un análisis básico de los datos correspondientes a las carreras del Departamento de Matemáticas. Para dicho análisis utilizamos series de tiempo. Con la función `ts()` de *R*, convertimos los datos del número total de alumnos, en una serie de tiempo. En la serie hay un dato para cada semestre del 2008-1 al 2020-1. Aplicamos la función `decompose()` a la serie de tiempo creada. Esta función utiliza el método de promedios móviles para descomponer la serie. Con ésto, obtuvimos un objeto de la clase *decomposed.ts* de *R*. A este objeto lo llamamos *num_total_alum.Comp*. Los elementos que tiene *num_total_alum.Comp* son los siguientes:

- *x:* Los valores observados de la serie de tiempo (X_t).

- *seasonal:* Valores estimados de la componente estacional de la serie de tiempo ($\hat{s}_t$).

- *figure:* Vector con los promedios del efecto estacional. La longitud del vector es igual a la frecuencia de los datos en la serie de tiempo. En este caso la longitud es 2 porque los datos son semestrales.

- *trend:* Valores estimados de la componente de tendencia $(\hat{m}_t)$.

- *random:* Valores estimados de la componente irregular $(\hat{y}_t)$.

- *type:* Tipo de variación estacional *("additive")*.

Graficamos *num_total_alum.Comp* para poder ver las componentes de la serie de tiempo (ver Figura 3.1). Se observan 4 diferentes gráficas, en la primera, de arriba hacia abajo, se observan los datos reales del número total de alumnos para cada semestre (X_t). En la segunda se muestra $\hat{m}_t$, la cual notamos que es creciente. En la tercera vemos $\hat{s}_t$ que nos indica que los datos tienen una estacionalidad semestral. En la cuarta se ve $\hat{y}_t$, la cual ya no tiene estacionalidad ni tendencia.

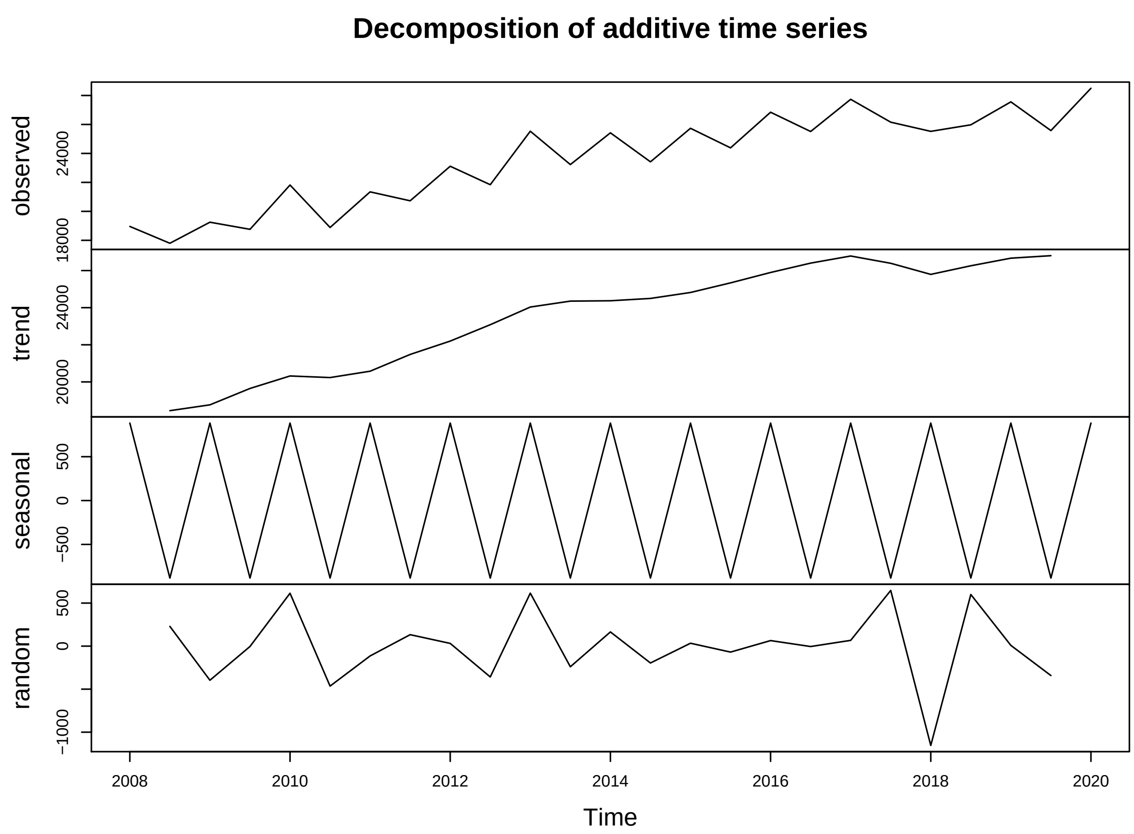


Figura 3.1: *Se muestra la descomposición por el método de promedios móviles con la función "decompose()". Los datos corresponden al número total de alumnos por semestre.*

Las técnicas de suavizamiento de series de tiempo son útiles para mostrar patrones subyacentes en los datos de las series de tiempo. El método que vamos a utilizar para mostrar dichos patrones de los datos es el método Holt-Winters aditivo. Este método se utiliza para describir y predecir valores con series de tiempo que tienen componentes de tendencia lineal y de estacionalidad.

Para probar éstos supuestos, existen diversas pruebas estadísticas. En las siguientes subsecciones veremos algunas de ellas. En cada una de las subsecciones presentaremos algunas gráficas de series de tiempo y otras de sus valores acumulados. Con ellas observaremos el comportamiento de los datos. Con ésto comprobaremos que los datos cumplen con los supuestos del método.

3.1.1. Prueba de tendencia

Al inicio de este capítulo vimos que se le llama tendencia al cambio, a largo plazo, del promedio de los datos. En la Figura 3.2 se muestran las gráficas del promedio del número de alumnos que toman clases por semestre de todas las materias. En la subfigura izquierda los datos están graficados como serie de tiempo. En la subfigura derecha la línea roja representa el ajuste de la tendencia, con un modelo de regresión lineal.

Observamos que los valores tienen una tendencia creciente, ésto nos indica que cada semestre, en promedio, el número de alumnos incrementa en la Facultad.

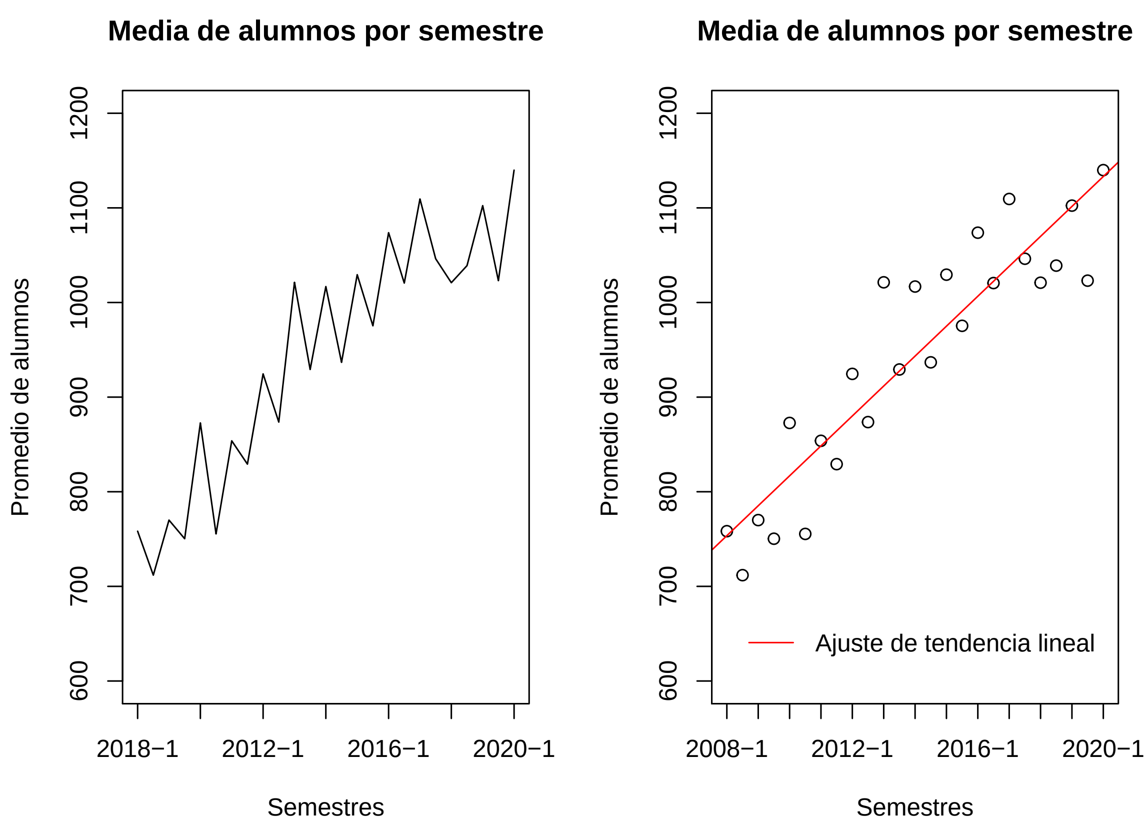


Figura 3.2: *Se observa una tendencia creciente en la media de alumnos por semestre. La información corresponde a los semestres del 2008-1 al 2020-1.*

Para probar que los datos no son aleatorios utilizamos la función `cox.stuart.test(X)`, de *R*. Dicha función corresponde a la prueba Cox-Stuart, que tiene como hipótesis nula H_0 : Los datos provienen de una muestra aleatoria. En la Figura 3.3 se muestran los resultados de la prueba.

```
> cox.stuart.test(vec_prom_total_alum)

        Cox Stuart test

data:  vec_prom_total_alum
statistic = 12, n = 12, p-value = 0.0004883
alternative hypothesis: non randomness
```

Figura 3.3: *En esta figura se muestran los resultados de la prueba Cox-Stuart. Esta prueba se utiliza para probar la aleatoriedad de los datos.*

Por [2] sabemos que se rechaza H_0 si *p-value* $\leqslant \alpha$, siendo α el nivel de significancia. Sea $\alpha = 0.01$. Como vemos en la Figura 3.3, *p-value* $= 0.0004 \leqslant 0.01 = \alpha$, por lo tanto se rechaza la hipótesis nula. Con ésto podemos concluir que los datos no provienen de una muestra aleatoria. Ésto nos indica que los datos pueden tener una tendencia creciente o decreciente.

Para probar que los datos tienen una tendencia creciente utilizamos la misma prueba pero con otra alternativa. El comando en *R* es: `cox.stuart.test(X,alternative="left.sided")`. Dicha prueba tiene como hipótesis nula H_0 : Los datos tienen una tendencia creciente.

En la Figura 3.4 se muestran los resultados de la prueba. Podemos ver que *p-value* $= 1 > 0.01 = \alpha$ por lo tanto no se rechaza H_0. Con ello concluimos que los datos tienen una tendencia creciente.

```
> cox.stuart.test(vec_prom_total_alum,alternative = "left.sided")

        Cox Stuart test

data:  vec_prom_total_alum
statistic = 12, n = 12, p-value = 1
alternative hypothesis: decreasing trend
```

Figura 3.4: *En esta figura se muestran los resultados de la prueba Cox-Stuart para tendencia. Con la alternativa elegida, esta prueba se utiliza para probar si los datos tienen una tendencia creciente.*

Finalmente la conclusión a la que llegamos con estas pruebas es que los datos tienen una tendencia lineal creciente.

3.1.2. Prueba de estacionalidad

En la Figura 3.5 se muestra la gráfica de barras con el número total de alumnos por semestre. A simple vista notamos que los datos tienen estacionalidad semestral. Podemos ver también que, en general, el número de alumnos de los semestres impares es mayor al de su siguiente semestre par. Este fenómeno los vimos en la Figura 1.1 al hacer el análisis correspondiente a los datos de la materia *Probabilidad I*.

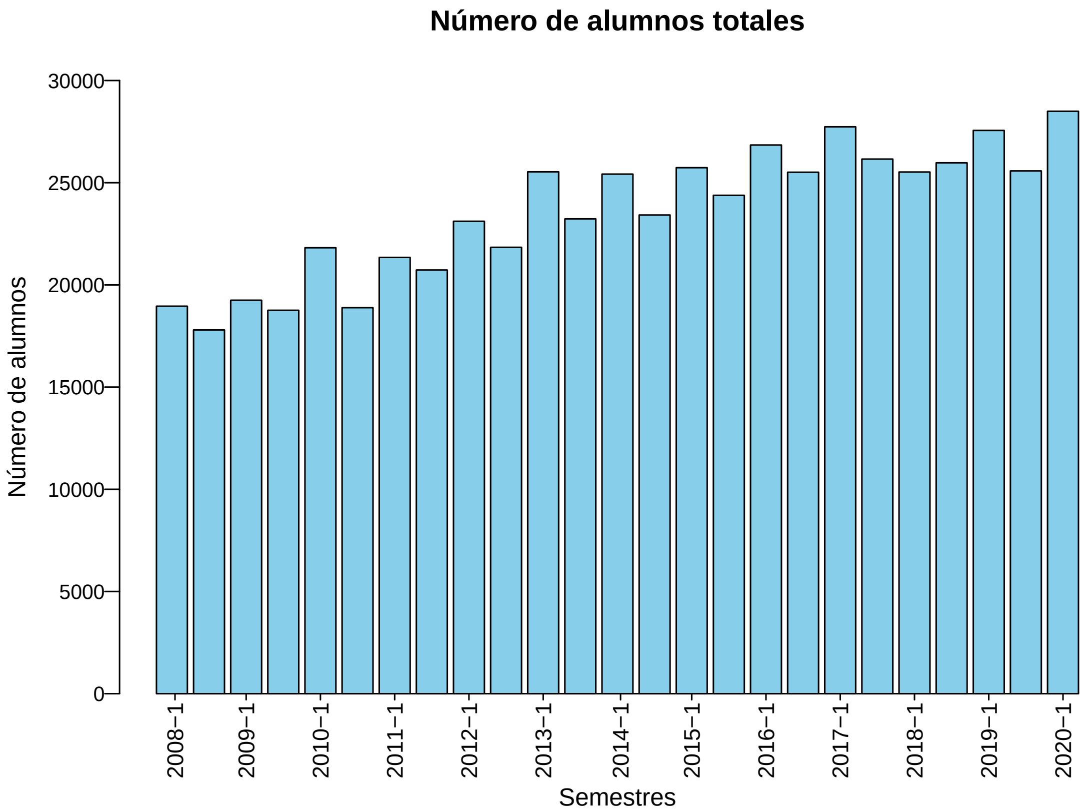


Figura 3.5: *En esta figura se muestra la gráfica de barras del número total de alumnos inscritos por cada semestre. Se observa que año con año el número aumenta. En general, el número de alumnos de los semestres impares es mayor que el de su respectivo semestre par.*

Para probar que los datos tienen variación estacional utilizamos la función `qs(X)`, de *R*. Dicha función corresponde a la prueba QS que tiene como hipótesis nula H_0 : No hay estacionalidad en la serie de tiempo. En la Figura 3.6 se muestran los resultados de la prueba QS. Podemos ver que *p-value* $= 1.47 \times 10^{-06} \leqslant 0.01 = \alpha$ por lo tanto se rechaza H_0. Con ello concluimos que los datos tienen variación estacional.

```
> qs(num_total_alum.ts,freq = 2)
Test used:  QS

Test statistic:  26.86
P-value:  1.473075e-06
```

Figura 3.6: *En esta figura se muestran los resultados de la prueba QS. Esta prueba se utiliza para probar si los datos tienen estacionalidad.*

3.1.3. Prueba de homocedasticidad

El término homocedasticidad se utiliza cuando algo tiene varianza constante. En nuestro caso, nos interesa probar que los datos con el número de alumnos totales tiene varianza constante.

Ésto para comprobar que el modelo de estacionalidad adecuado para nuestros datos es el aditivo.

En la Figura 3.7 se muestra la gráfica de la desviación estándar del número de alumnos por grupo y por semestre de todas las materias. Observamos que los valores se mantienen constantes a lo largo del tiempo, en un rango de entre 24 y 29 alumnos.

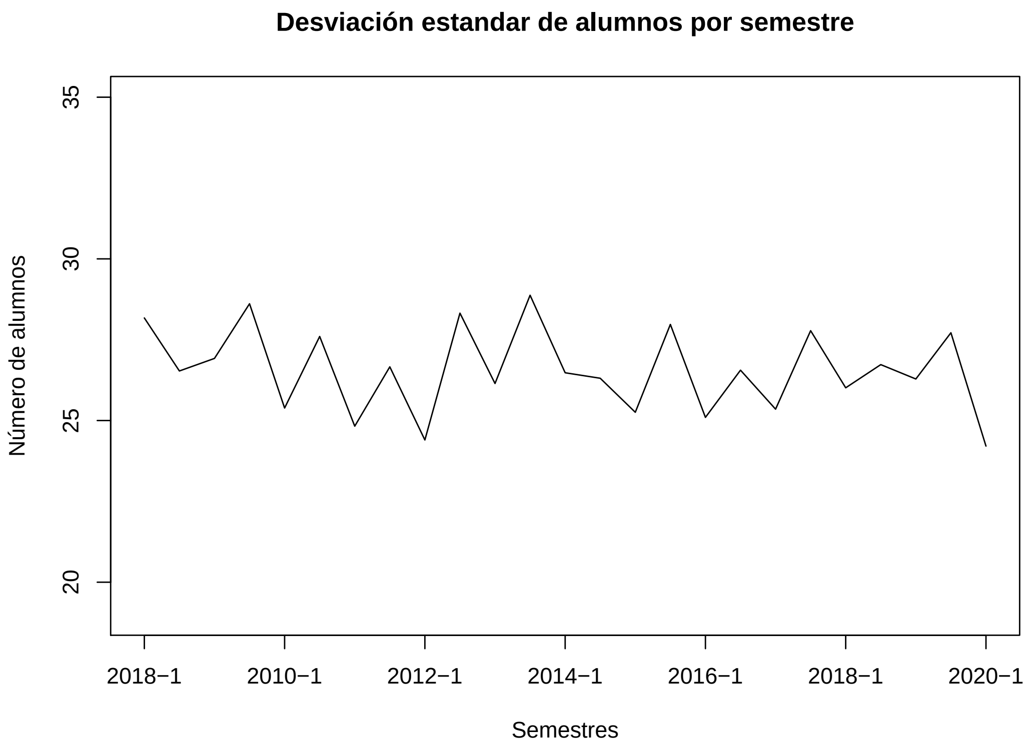


Figura 3.7: *Se muestra la desviación estándar del número de alumnos por semestre. Se puede ver que el comportamiento de los datos es constante a lo largo del tiempo.*

Utilizamos la prueba Breusch-Pagan que tiene como supuesto que los datos tienen una distribución normal. Para probar que los datos se distribuyen normal, utilizamos la prueba Jarque-Bera. El comando en *R* es: `jarque.bera.test(X)`. Dicha prueba tiene como hipótesis nula H_0 : Los datos provienen de una distribución normal.

En la Figura 3.8 vemos el resultado de la prueba Jarque-Bera. Notamos que *p-value* $= 0.40 > 0.01 = \alpha$ por lo tanto no se rechaza H_0, entonces la distribución de los datos es normal.

```
> jarque.bera.test(num_total_alum.ts)

        Jarque Bera Test

data:  num_total_alum.ts
X-squared = 1.791, df = 2, p-value = 0.4084
```

Figura 3.8: *Se muestran los resultados de la prueba Jarque-Bera. Esta prueba se utiliza para probar si los datos tienen una distribución normal.*

Para probar la homocedasticidad de los datos, utilizamos la función `bptest(lm(X~t))`. Esta función corresponde a la prueba Breusch-Pagan. El ajuste del modelo lineal con la función `lm(X~t)` es con respecto al tiempo. La prueba mencionada tiene como hipótesis nula H_0 : La varianza es constante.

En la Figura 3.9 Se muestra el resultado de la prueba. Vemos que $p\text{-}value = 0.82 > 0.01 = \alpha$ por lo tanto no se rechaza H_0, entonces la varianza de los datos es constante.

```
> bptest(lm(num_total_alum.ts~t))

        studentized Breusch-Pagan test

data:  lm(num_total_alum.ts ~ t)
BP = 0.050991, df = 1, p-value = 0.8213
```

Figura 3.9: *Se muestran los resultados de la prueba Breusch-Pagan. Esta prueba se utiliza para probar si los datos tienen varianza constante.*

Con las pruebas de tendencia y de estacionalidad confirmamos que se puede utilizar el método Holt-Winters. La prueba de homocedasticidad nos ayuda a verificar que el modelo de estacionalidad que debemos utilizar es el aditivo. Con estas observaciones concluímos que el método Holt-Winters aditivo es el método adecuado para poder hacer predicciones con nuestros datos.

3.2. Análisis estadístico por grupo de datos

En la Figura 3.10 vemos la gráfica del número de alumnos separado por semestres pares e impares. Vemos con mayor claridad lo que ocurre en la Figura 3.5, los datos efectivamente tienen una tendencia creciente. Notamos que el número de alumnos de los semestres impares es mayor al número total de alumnos de los semestres pares en todos los semestres, salvo en el semestre 2018-1 en donde el número de alumnos es menor a los de los semestres adyacentes.

En la Figura 3.11 observamos los dos histogramas con el número total de alumnos de semestres pares e impares con sus respectivas densidades ajustadas. Notamos que hay una ligera diferencia entre el número de alumnos de los semestres pares con respecto al número de alumnos de los semestres impares.

Existe una mayor cantidad de grupos en los semestres pares con un menor número de alumnos, que en los semestres impares. Hay una mayor cantidad de grupos en los semestres impares contra los semestres pares, que tienen entre 35 y 100 alumnos. Tanto para los semestres pares como para los impares, el comportamiento de las densidades ajustadas es muy parecido.

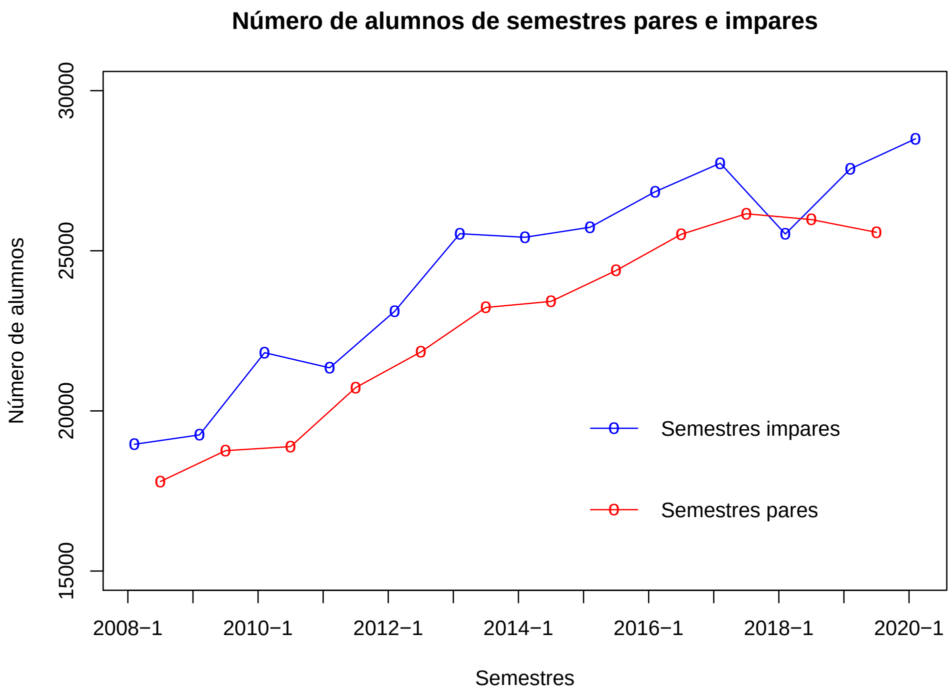


Figura 3.10: *Se muestran las series de tiempo del número de alumnos en semestres pares e impares. Se observa una tendencia creciente y en general el número de alumnos de semestres impares (línea azul) es mayor al número de alumnos de semestres pares (línea roja).*

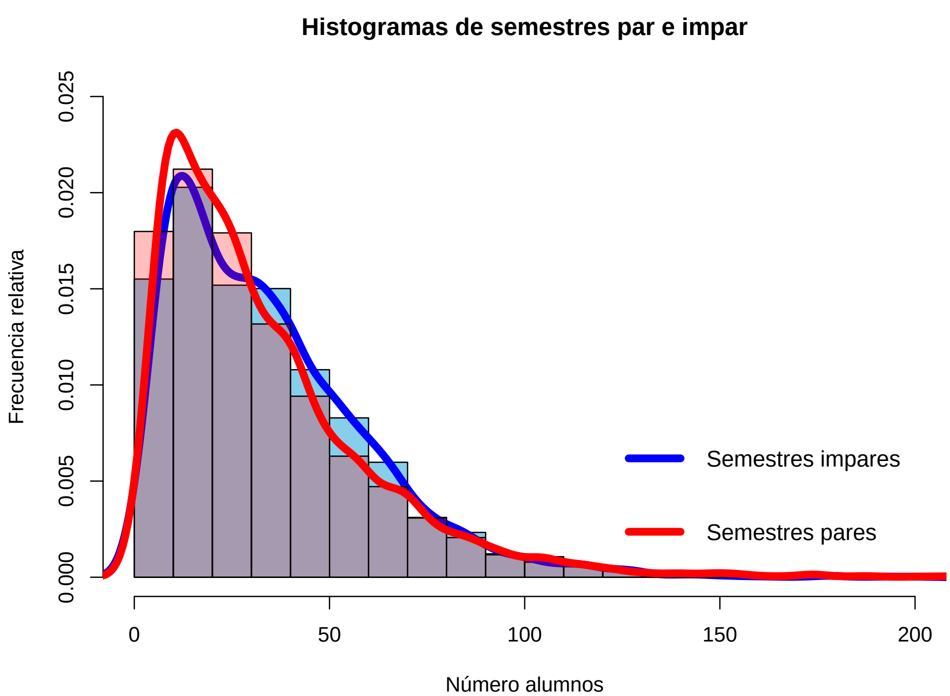


Figura 3.11: *Se muestran los histogramas con el número de alumnos en semestres pares e impares. Las densidades ajustadas son muy parecidas, no importa si los datos son de semestres pares o impares.*

En la Figura 3.12 mostramos la gráfica del número de alumnos por turno: matutino y vespertino. Se puede observar que en todo momento el número de alumnos del turno matutino es mayor al número de alumnos del turno vespertino.

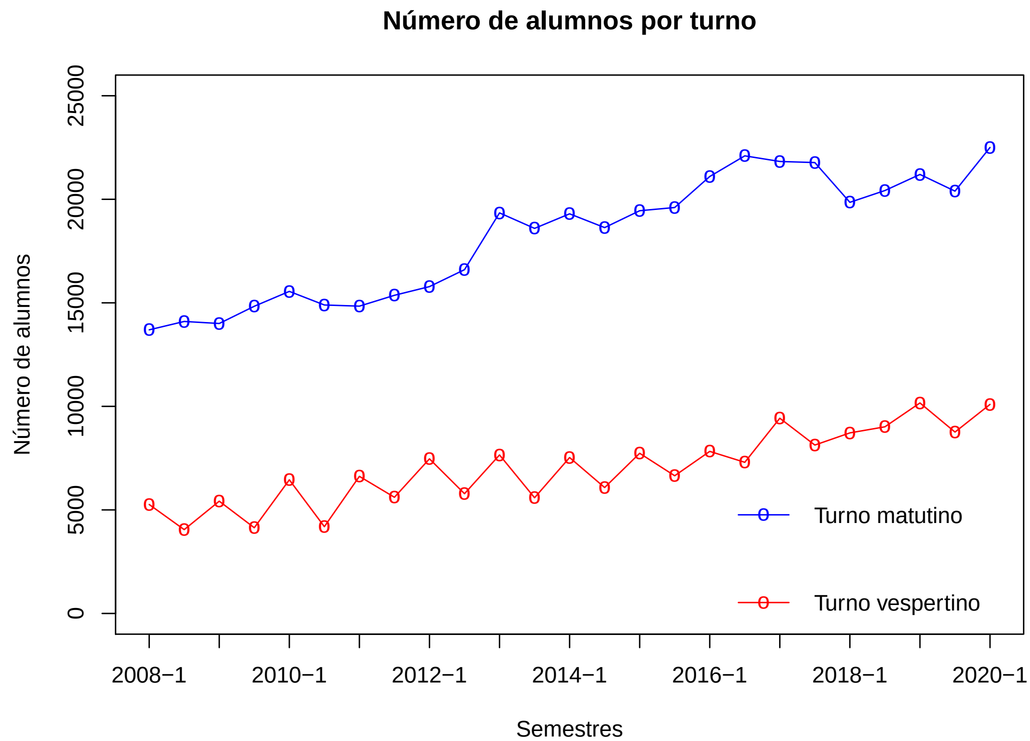


Figura 3.12: *Se muestran las series de tiempo del número de alumnos por turno de todos los semestres. Se observa que el número de alumnos del turno matutino (línea azul) es siempres mayor al número de alumnos del turno vespertino (línea roja).*

Los datos que se graficaron en los histogramas de la Figura 3.13 corresponden al número de alumnos totales por turno (matutino y vespertino) de cada semestre. Notamos que las densidades ajustadas de cada turno son completamente diferentes. Al ver la gráfica podemos concluir que hay más alumnos en el turno matutino que en el vespertino.

3.3. Análisis estadístico por carrera

Es importante recordar que dentro de las carreras existe un tronco común. Es decir, comparten muchas de las materias impartidas en los primeros 4 semestres, por lo que muchos de los grupos de una carrera se encuentran en otra. Cabe mencionar que el número máximo de alumnos por grupo para la carrera de Ciencias de la Computación es 211 y para las otras carreras es 353.

En la Figura 3.14 vemos cuatro histogramas con el número de alumnos por grupo, uno para cada carrera del Departamento de Matemáticas. La escala del eje Y es igual para todos los histogramas. De esta manera podemos observar que en las carreras de Actuaría, Ciencias de la Computación y Matemáticas, se tiene la mayor concentración en los grupos de entre 10 y

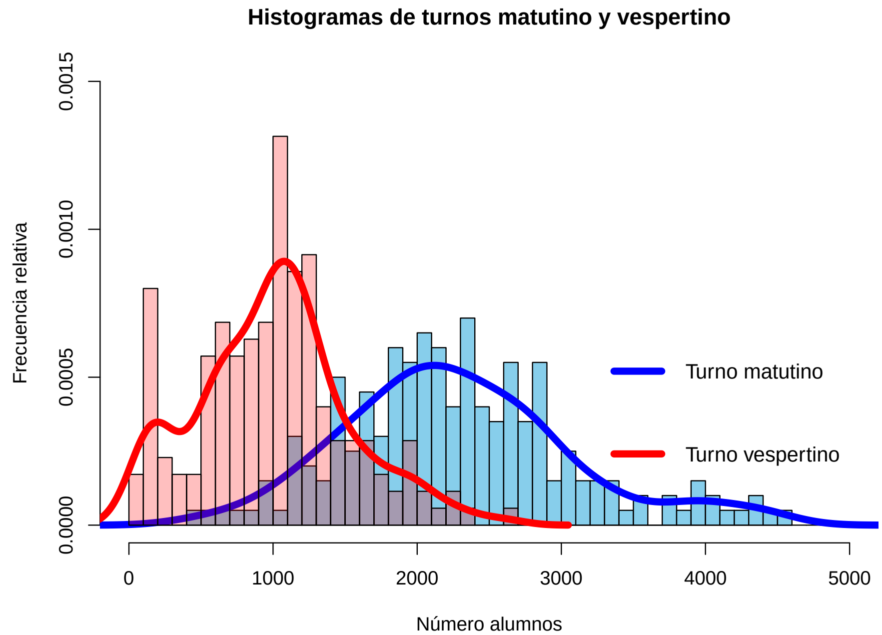


Figura 3.13: *Se muestran los histogramas del número de alumnos en los turnos matutino y vespertino. Se puede concluir que hay más alumnos en el turno matutino que en el vespertino. Sus densidades ajustadas son diferentes.*

20 alumnos. La carrera de Matemáticas Aplicadas tiene su mayor concentración en los grupos que tienen entre 20 y 30 alumnos.

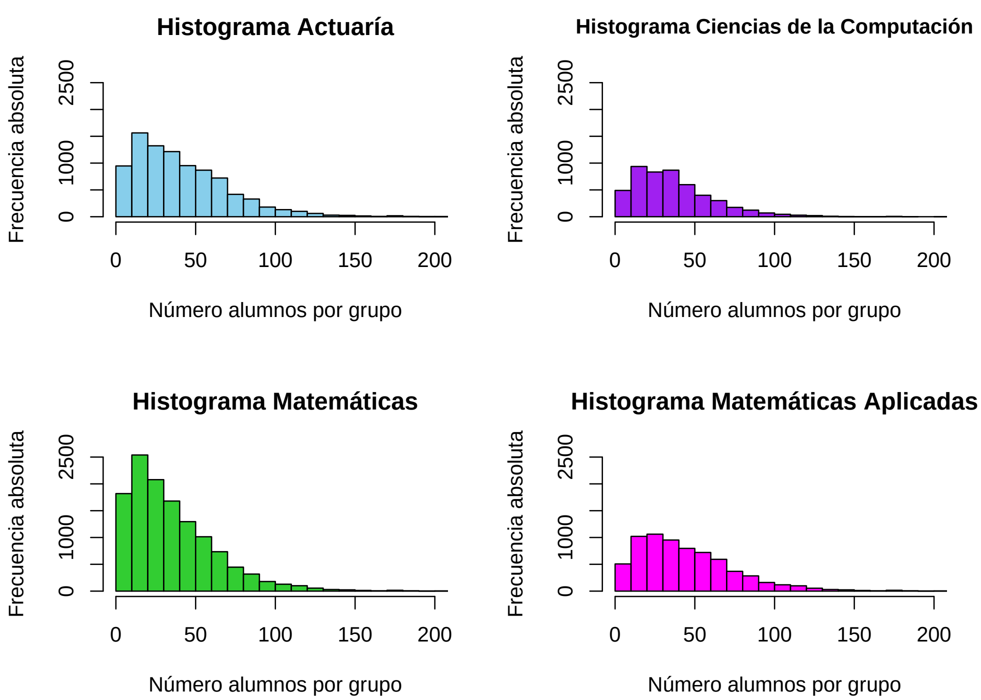


Figura 3.14: *Se muestran los histogramas con el número de alumnos por grupo. Hay un histograma para cada carrera del Departamento de Matemáticas.*

En la Figura 3.15 vemos una gráfica con las densidades ajustadas a los datos del número de alumnos por grupos para cada carrera. Al ver la densidad ajustada a los datos de la carrera de Matemáticas vemos que tiene una mayor concentración de grupos que tienen aproximadamente entre 10 y 30 alumnos, a diferencia de las otras carreras.

También podemos observar que en Actuaría y en Matemáticas Aplicadas hay una mayor concentración en los grupos que tienen aproximadamente entre 50 y 75 alumnos, que en Matemáticas o en Ciencias de la Computación.

Si vemos la densidad ajustada a los datos de Ciencias de la Computación notamos que hay dos grandes concentraciones, una en los grupos de aproximadamente entre 20 y 30 alumnos y otra entre 40 y 50 alumnos. Con esta gráfica podemos ver con mayor claridad lo que observamos en la Figura 3.14, el comportamiento es similar para todas las carreras pero cada una tiene su propia distribución.

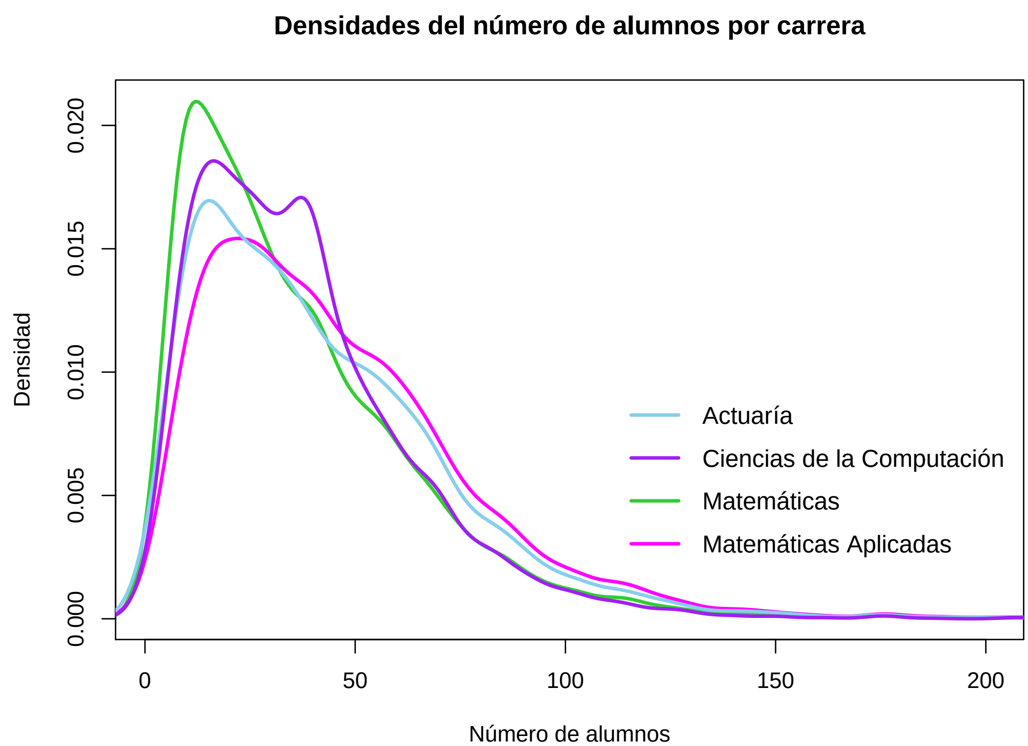


Figura 3.15: *Se muestran las densidades ajustadas del número de alumnos por carrera del Departamento de Matemáticas.*

3.4. Distribución del tamaño de los grupos

En la Figura 3.16 se muestra el histograma del número de alumnos por grupo de todos los semestres, desde el 2008-1 hasta el 2020-1. Observamos el mismo comportamiento que en las Figuras 3.11, 3.14 y 3.15. La mayor frecuencia se encuentra en los grupos que tienen entre 10 y 20 alumnos.

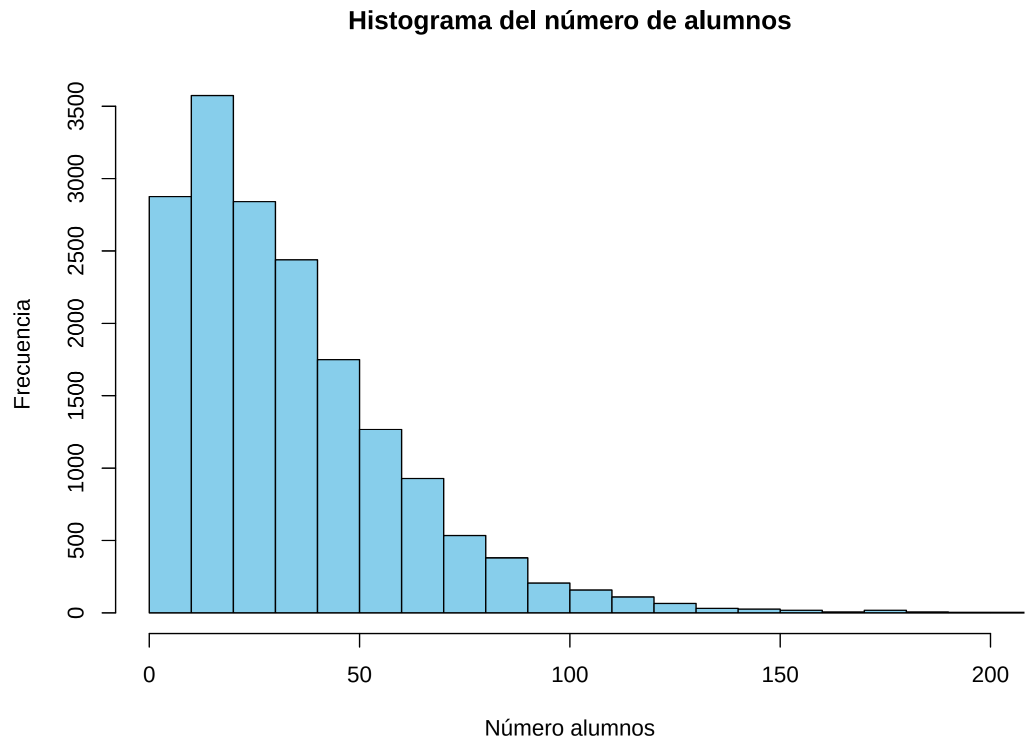


Figura 3.16: *Se muestra el histograma del número de alumnos por grupo de todos los semestres. La información es de los semestres del 2008-1 al 2020-1. Vemos una mayor concentración en los grupos que tienen entre 10 y 40 alumnos.*

En la Figura 3.17 vemos diferentes líneas con las densidades ajustadas a los valores del número de alumnos por grupo de cada semestre. Cada línea corresponde a un semestre. Se tomaron los datos de 25 semestres, del 2008-1 al 2020-1.

En dicha figura, las líneas de color verde corresponden a las densidades ajustadas a los datos de los semestres del 2008-1 al 2012-2. Las de color rosa corresponden a los semestres del 2013-1 al 2017-2. Las de color azul corresponden a los semestres 2018-1 al 2020-1. Notamos que el comportamiento va cambiando conforme pasa el tiempo.

Vemos que en los semestres más antiguos se tiene una concentración mayor en los grupos que tienen aproximadamente entre 10 y 30 alumnos. En los semestres más recientes la mayor concentración se tiene en los grupos con aproximadamente entre 25 y 50 alumnos. Esto lo podemos explicar con el hecho de que cada semestre incrementa el número de alumnos inscritos en la facultad, por lo tanto el tamaño de los grupos aumenta.

También podemos observar que conforme pasa el tiempo la media y la varianza aumentan. Es decir, en semestres antiguos se tiene una menor media y varianza. Ésto comparado con los semestres más actuales, los cuales tienen una media y varianza mayor.

En los semestres del 2013-1 al 2017-2 se tuvieron grupos con más de 350 alumnos. En los semestres más recientes el número máximo de alumnos por grupo fue alrededor de 250. Ésto lo podemos explicar por las medidas que se tomaron después del sismo del 19 de septiembre de 2017, con respecto al tamaño de los grupos. El número de alumnos inscritos no puede ser mayor al número de lugares disponibles por salón.

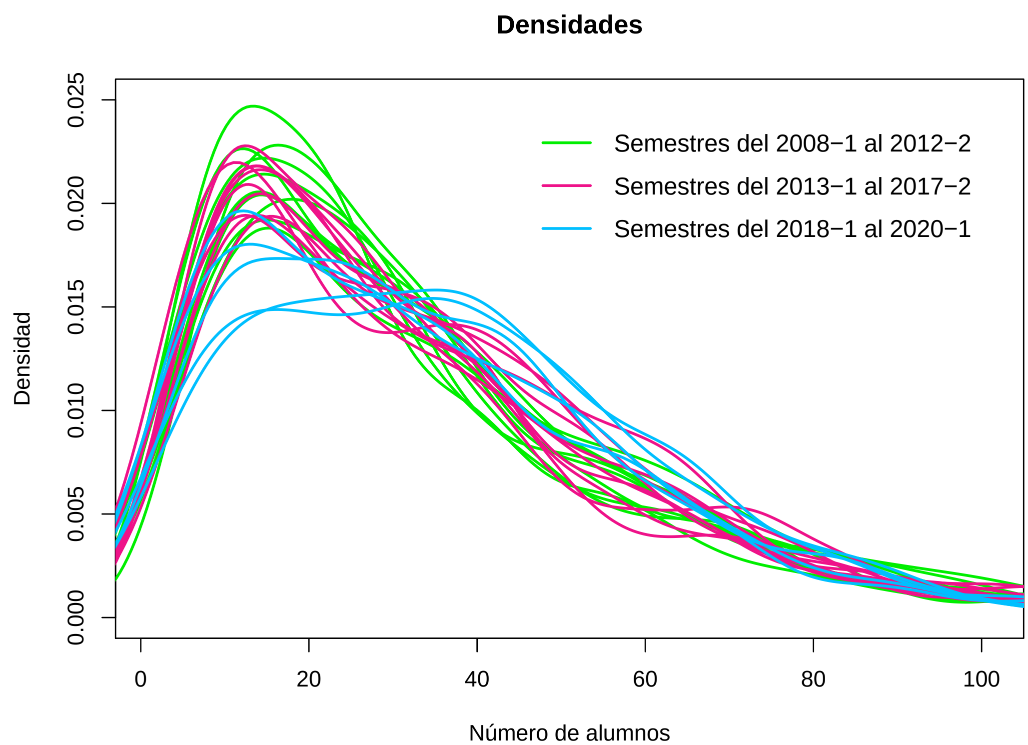


Figura 3.17: *Se muestran las densidades ajustadas del número de alumnos por grupo de cada semestre. Cada línea corresponde a un semestre entre el 2008-1 y el 2020-1.*

Viendo las Figuras 3.16 y 3.17, podríamos concluir que la distribución que mejor se ajusta al tamaño de los grupos es la distribución Poisson por la forma en la que están distribuidos los datos. Para probar esta hipótesis utilizamos la función `ks.test(X,Y)`, de *R*, para hacer la prueba de Kolmogorov-Smirnov.

La prueba de Kolmogorov-Smirnov, dice que se rechaza H_0 cuando $D_n > D_n^{1-\alpha}$. Donde $D_n^{1-\alpha}$ nos indica el valor en donde inicia la región de rechazo para un nivel de significancia de α y n es el número de datos de la muestra. Tomamos como hipótesis nula $H_0 : X$ y Y tienen la misma distribución.

Definimos a X como el vector con el número de alumnos por cada grupo del semestre 2008-1 al 2020-1. Definimos a Y como un vector de números aleatorios de una distribución $Poisson(\lambda)$. Por el resultado B.1, del apéndice B, sabemos que el estimador máximo verosímil de λ para la distribución $Poisson(\lambda)$ es la media de los datos. Con este estimador $(\hat{\lambda} = 34.18)$, obtuvimos los números aleatorios de Y. Utilizamos el siguiente comando en *R*: `Y<-rpois(1000,34.18)`. Tenemos que $Y \sim Poisson(34.18)$.

Por [10] sabemos que:

$$D_n^{1-\alpha} = \sqrt{\frac{ln\left(\frac{1}{\alpha}\right)}{2n}} - 1.6693n^{-1} - 0.20562n^{-\frac{3}{2}} \tag{3.5}$$

En nuestro caso los valores de las variables son: $n = 17{,}246$ y $\alpha = 0.01$. Sustituyendo en la ecuación 3.5 tenemos que $D_{17246}^{0.99} = 0.01$. Con la función `ks.test(X,Y)`, de *R*, obtenemos el valor de $D_{17246} = 0.39$.

Como $D_{17246} = 0.39 > 0.01 = D_{17246}^{0.99}$, entonces se rechaza H_0, por lo tanto los datos no siguen una distribución Poisson con $\lambda = 34.18$.

Hicimos otra prueba suponiendo que los datos tienen una distribución $Normal(\mu,\sigma)$. Para simular los datos de Y utilizamos los estimadores máximo verosímiles de μ y σ. Estos estimadores los obtuvimos con la función `fitdistr(X, densfun="normal")`, en *R*. Los valores de los estimadores son $\hat{\mu} = 34.18$ y $\hat{\sigma} = 26.57$. El resultado de la función de la prueba de Kolmogorov-Smirnov es $D_{17246} = 0.10$.

Como $D_{17246} = 0.10 > 0.01 = D_{17246}^{0.99}$, entonces se rechaza H_0, por lo tanto los datos no siguen una distribución $Normal(34.18, 26.57)$.

Hicimos más pruebas con otras distribuciones y en todos los casos rechazamos la hipótesis nula. En la Figura 3.18 vemos únicamente los casos que expusimos en esta sección.

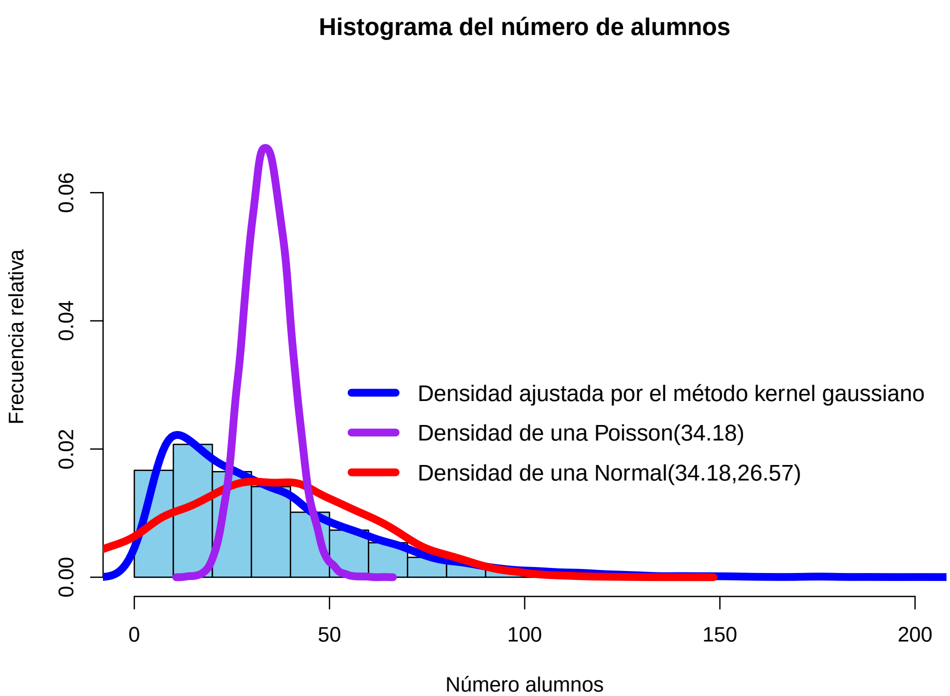


Figura 3.18: *Se muestran 3 densidades ajustadas. La línea azul es la ajustada con la función "density()" en R. La morada es de una Poisson(34.18). La roja de una Normal(34.18,26.57). Ninguna de las distribuciones propuestas se ajustan de manera adecuada a los datos.*

El histograma de la Figura 3.18 representa las frecuencias relativas del número de alumnos por grupo para cada semestre. La línea azul es la densidad ajustada con el método kernel gaussiano, generada por la función `density()`, en *R*. La línea morada es la densidad de n números aleatorios con distribución $Poisson(34.18)$. La línea roja es la densidad de n números aleatorios con distribución $Normal(34.18, 26.57)$.

Hicimos más pruebas con otras distribuciones y en todos los casos rechazamos la hipótesis nula. Con estos resultados concluímos que el ajuste que se pudiera hacer a los datos tendría que ser en dos o más partes. Es decir, tener 2 o más distribuciones para el mismo conjunto de datos en diferentes secciones. Un ejemplo de una posible partición de los datos es que se puede ajustar una distribución para los datos que están entre 0 y 100, y otra distribución para los datos mayores a 100. Esto debido a que a pesar de ser pocos los datos mayores a 100, si impactan en la distribución total. Un análisis con esta propuesta se puede ver en la Sección 4.7.

3.5. Comportamientos por hora

En esta sección veremos algunas gráficas cuyo eje x corresponde a las horas en las que se imparten las clases. Se empieza por la clase de 7-8hrs y se termina con la clase de 21-22hrs. Primero mostraremos el comportamiento del promedio de grupos por hora y después el comportamiento del promedio del número de alumnos por hora.

En la Figura 3.19 vemos la gráfica de barras con el número promedio de grupos por hora. Tomamos la información de 25 semestres. Observamos una disminución considerable del número de grupos a las 15hrs. Podemos concluir que es debido a que a esa hora, usualmente la gente sale a comer. A las 21hrs se tiene el menor número de grupos, esto se puede explicar por el hecho de que es la última clase impartida en la Facultad.

Hay un descenso leve a las 9hrs donde se pudiera suponer que la gente sale a desayunar. Desde las 7hrs se pueden encontrar clases como *Cálculo Diferencial e Integral*. Pero en general, las clases impartidas a las 7hrs y a las 8hrs suelen ser materias exclusivas para los actuarios como *Teoría del Seguro, Matemáticas Actuariales del Seguro de Personas I y II* o *Matemáticas Actuariales para Seguro de Daños, Fianzas y Reaseguro*. Podemos decir que a partir de las 9 de la mañana se imparten materias de todas las carreras.

A las 10 de la mañana se tiene el número máximo de grupos. Con esta información se podría medir la capacidad que debería de tener la Facultad en cuanto al número de salones necesarios para cubrir la demanda de grupos. Si se está preparado para cubrir la demanda del pico más alto de todas las horas, entonces los demás casos están cubiertos por tener un menor número de grupos.

En la Figura 3.20 se muestra la gráfica de barras con el promedio del número de alumnos por hora. Notamos que el comportamiento de ésta gráfica es muy similar al de la gráfica mostrada en la Figura 3.19. El pico más alto de los datos también se tiene a las 10 de la mañana y el menor número de alumnos se encuentra a las 21hrs. También hay una disminución considerable a las 15hrs.

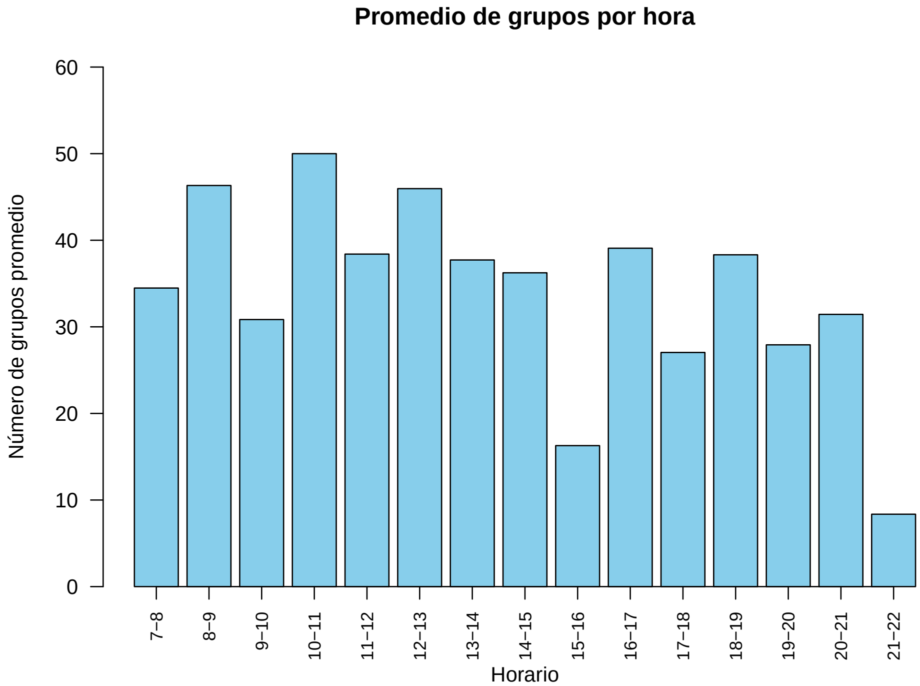


Figura 3.19: *Se muestra la gráfica de barras del número promedio de grupos por hora. Se observa una disminución considerable a las 15hrs y a las 21hrs. El valor más alto se encuentra a las 10hrs.*

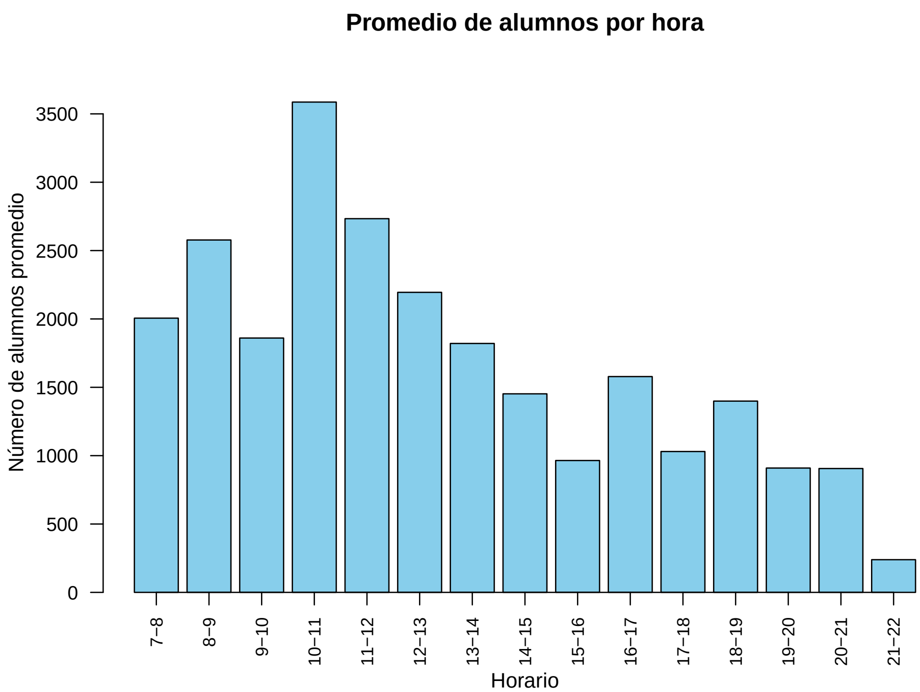


Figura 3.20: *Se muestra la gráfica de barras del número promedio de alumnos por hora. Notamos una disminución de los valores a las 9hrs, a las 15hrs y a las 21hrs. El valor más alto lo encontramos a las 10hrs.*

Viendo las Figuras 3.19 y 3.20 podemos concluir que existe una correlación entre el número promedio de grupos por hora y el número promedio de alumnos por hora. Por ejemplo, si no hay alumnos que tomen clases a las 15hrs entonces no tiene caso que se abran grupos a esa hora. Análogamente para las 21hrs. Por el contrario entre más alumnos haya por hora, se deben abrir más grupos a esas horas, como es el caso de las 10hrs.

Capítulo 4

Simulación

La simulación es un proceso que nos permite estudiar el comportamiento de un sistema complejo y difícil de examinar de manera analítica. Nos ayuda a determinar de manera empírica las probabilidades de ciertos eventos. También nos permite experimentar con diversos supuestos que podrían ser muy costosos o riesgosos de realizar físicamente, como enseñar a los pilotos a volar un avión.

Algunas áreas de aplicación son: biología, estadística, medicina, química, matemáticas, investigación de operaciones, física y ciencias sociales. Los ejemplos de su aplicación van desde simular el lanzamiento de una moneda justa, hasta la simulación de colisiones de átomos en un acelerador de partículas.

Actualmente se combinan diferentes metodologías de simulación con el software disponible, el análisis de sensibilidad y la optimización estocástica. Ésto para obtener un mejor resultado al momento de simular sistemas que son cada vez más complejos como las redes neuronales.

En este trabajo utilizamos la simulación para poder realizar predicciones en base a datos históricos. Tomamos la información de los horarios de la Facultad y con ellos simulamos la demanda del número de alumnos para el siguiente semestre. Con esta demanda hicimos los esqueletos necesarios para realizar la asignación de horarios.

En *R* realizamos la función `gen_asignacion()` encargada de generar la asignación de horarios, materias y profesores. En la Figura 4.1 se muestra el diagrama de flujo que sigue dicha función. A lo largo de este capítulo explicaremos los pasos (3)-(9) mostrados en el diagrama. Cabe aclarar que los pasos (1) y (2) corresponden a las Secciones 2.2 y 2.4, respectivamente.

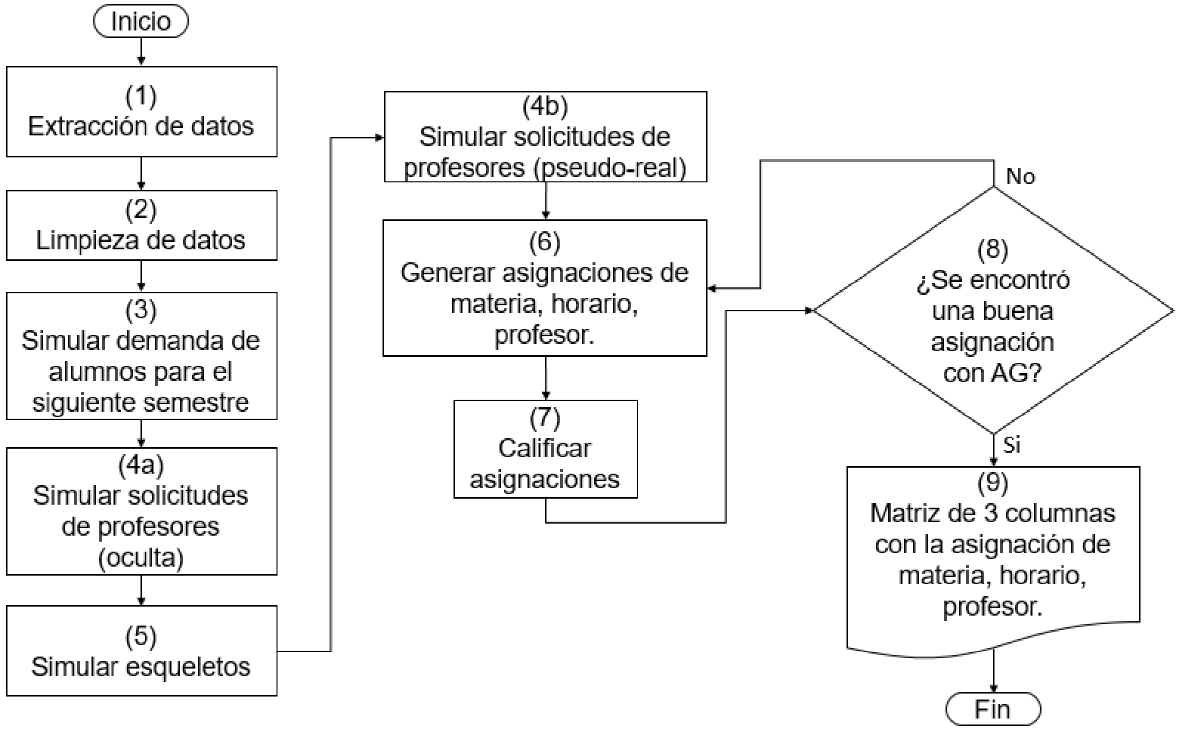


Figura 4.1: *Se muestra el diagrama de flujo de la función "gen_asignacion()". Se pueden ver los pasos para obtener la matriz de asignaciones.*

4.1. Obtención de nombres de materias

Antes de iniciar las simulaciones primero obtuvimos el vector *vec_nom_materias_total*. Éste contiene el nombre de las materias encontradas en la matriz *m_grande_total* de los semestres 2008-1 al 2020-1. Dicho vector no tiene nombres repetidos y contiene $m = 202$ materias.

En esta sección vamos a explicar cómo obtuvimos los m nombres de las materias que vamos a utilizar. El motivo de obtener el vector *vec_nom_materias_total* antes de hacer las simulaciones es para evitar problemas como el que vimos en la Figura 2.11, de repetición de información.

Primero obtuvimos un vector con los nombres de las materias en la matriz *m_grande_total*. Aplicamos la función `unique()` de *R* y obtuvimos un vector de 333 materias. En este vector se pueden encontrar nombres repetidos y nombres de materias que ya no se imparten. Definimos la matriz *mat_nom_materias_total* la cual tiene 22 columnas:

- La primer columna contiene el nombre que vamos a utilizar para las simulaciones y para las asignaciones. En la mayoría de los casos elegimos el nombre más reciente de la materia. Cabe aclarar que hubo algunos casos que elegimos el nombre que lleva la materia en la carrera de Actuaría en lugar del más reciente.

- La segunda columna contiene el número de materia con respecto a la primer columna.

- Las columnas 3-22 contienen todos los posibles nombres asociados al nombre en la primer columna. Cabe aclarar que no todas estas columnas están llenas.

Revisamos caso por caso para no tener nombres repetidos. En el caso de los seminarios, los agrupamos de acuerdo a los posibles nombres que han tenido. Las materias optativas que ya no son impartidas las agrupamos en temas similares. Ésto último para conservar toda la información posible.

Finalmente las dimensiones de la matriz *mat_nom_materias_total* son 202×22. Con la primer columna de dicha matriz, obtuvimos el vector *vec_nom_materias_total*. Los nombres del vector son los que utilizaremos en las siguientes secciones para realizar las simulaciones y las asignaciones.

4.2. Obtención de los parámetros q_1 y q_2

En esta sección vamos a explicar cómo obtuvimos los valores de q_1 y q_2. Éstos son parámetros que se introducen en la función `hw()` de *R*. Dicha función corresponde al método Holt-Winters aditivo.

Los valores de q_1 y q_2 representan los cuantiles utilizados al calcular los intervalos de confianza. Por ejemplo, si $q_1 = 80$ entonces se calcula el intervalo al 80 % de confianza. Si se introducen a la función los dos parámetros entonces se calculan dos intervalos, uno al q_1 % de confianza y el otro al q_2 % de confianza.

Primero definimos los parámetros generales necesarios para las simulaciones:

1. Fijamos la semilla con el comando `set.seed(8654)`, en *R*.

2. Elegimos 3 semestres para simular la demanda del número de alumnos. Los seleccionamos de los semestres que ya teníamos guardados con información real. Hicimos una comparación entre nuestros datos simulados y los reales de cada semestre. Los semestres que elegimos fueron: 2019-1, 2019-2 y 2020-1.

3. Fijamos $k = 5$ (número de semestres que se tienen como ventana de información).

4. Fijamos $num\_sim = 10$ (número de simulaciones de la demanda de alumnos para el semestre a simular).

Después fijamos 5 materias que consideramos representativas para hacer las pruebas iniciales: *Cálculo Diferencial e Integral I*, *Demografía*, *Modelos no Paramétricos y de Regresión*, *Administración de Riesgos Financieros* y *Temas Selectos de Investigación de Operaciones*.

Tomamos 12 posibles combinaciones de valores para q_1 y q_2, las cuales podemos ver en la Tabla 4.1. La letra L indica que se tomó la cota inferior del intervalo al q_1 % de confianza. La letra U indica que se tomó la cota superior del intervalo al q_2 % de confianza.

Con estas cotas formamos intervalos de tipo (Lq_1, Uq_2). De éstos intervalos obtuvimos el número de alumnos simulados para los 3 diferentes semestres previamente definidos y para cada una de las 5 materias elegidas.

$q_1 \backslash q_2$	**80**	**85**	**90**	**99**
80	-	L80,U85	L80,U90	L80,U99
85	L85,U80	-	L85,U90	L85,U99
90	L90,U80	L90,U85	-	L90,U99
99	L99,U80	L99,U85	L99,U90	-

Tabla 4.1: *Tabla que muestra todas las combinaciones de los intervalos formados con las cotas inferiores y superiores de los intervalos de confianza al q_1 % y al q_2 %.*

Una vez hecha la simulación obtuvimos dos matrices:

1. Matriz de diferencias relativas: Esta matriz se genera al restar, los datos reales menos los simulados y después dividirlos entre los reales. Ésta operación se repite para cada materia y para cada simulación.

2. Matriz con información por materia: Esta matriz tiene 6 columnas: *materia, intervalo, mín, media, máx* y *sd*. En el renglón *i* se tienen los datos de la matriz de diferencias relativas de la i-ésima materia para el intervalo (Lq_1, Uq_2) correspondiente. Por ejemplo, en el primer renglón de la Figura 4.2 vemos que se utilizó el intervalo $(L80, U85)$ para obtener el número de alumnos simulados para el siguiente semestre de la materia *Cálculo Diferencial e Integral I*. Las columnas 3 y 5 corresponden al mínimo y al máximo error relativo de la materia mencionada. Las columnas 4 y 6 indican la media y la varianza de los errores relativos de todas las simulaciones hechas para *Cálculo Diferencial e Integral I*.

Materia	Intervalo	Min	Media	Max	sd
Cálculo Diferencial e Integral I	L80,U85	-2.622222	-0.21911543	0.8627586	0.7287619
Demografía	L80,U85	-1.985714	-0.09395821	0.8378378	0.4695739
Modelos no Paramétricos y de Regresión	L80,U85	-6.922222	-0.45848861	1.0000000	1.5742750
Administración de Riesgos Financieros	L80,U85	-1.816667	-0.03119518	0.6312500	0.3145900
Temas Selectos de Investigación de Operaciones	L80,U85	-2.300000	-0.05121275	0.9384615	0.4202550
Cálculo Diferencial e Integral I	L80,U90	-2.588889	-0.25311699	0.7220690	0.7108189
Demografía	L80,U90	-3.228571	-0.20226571	0.7270270	0.6654177
Modelos no Paramétricos y de Regresión	L80,U90	-6.744444	-0.48396359	1.0000000	1.6007042
Administración de Riesgos Financieros	L80,U90	-2.316667	-0.04418860	0.6375000	0.3924680
Temas Selectos de Investigación de Operaciones	L80,U90	-2.233333	-0.05595981	0.9461538	0.4210018

Figura 4.2: *Se muestran los primeros 10 renglones de la tabla obtenida con información de la matriz de diferencias relativas.*

Decidimos elegir q_1 y q_2 en base a la desviación estándar. A partir de la matriz con información por materia obtuvimos una matriz de dos columnas que se muestra en la Figura 4.3. La nueva matriz contiene en su primer columna el intervalo (Lq_1, Uq_2) correspondiente. En la segunda el promedio de la desviación estándar para cada intervalo de las 5 materias.

Intervalo	Promedio_sd
L85,U80	0.6877112
L90,U80	0.6893502
L80,U85	0.7014912
L90,U85	0.7218125
L80,U90	0.7580821
L85,U90	0.7705116
L99,U90	0.8014339
L90,U99	0.9032661
L99,U80	0.9045421
L99,U85	0.9422762
L85,U99	0.9579213
L80,U99	0.9615854

Figura 4.3: *Se muestra la tabla con el promedio de la desviación estándar para 5 materias y 12 diferentes intervalos.*

Los datos en la Figura 4.3 están ordenados de menor a mayor con respecto al promedio de la desviación estándar. Para la segunda prueba elegimos los primeros 6 intervalos de dicha tabla y seleccionamos otras 10 materias: *Álgebra Lineal I*, *Álgebra Superior II*, *Cómputo Evolutivo*, *Análisis Matemático IV*, *Matemáticas Actuariales para Seguro de Daños, Fianzas y Reaseguro*, *Análisis Numérico*, *Teoría de la Medida I*, *Introducción a las Matemáticas Discretas*, *Inglés I* y *Cálculo Diferencial e Integral IV*. La tabla con el promedio de la desviación estándar de sus datos se puede ver en la Figura 4.4.

Intervalo	Promedio_sd
L85,U90	0.4694684
L85,U80	0.4805732
L80,U90	0.4893892
L90,U85	0.4992955
L80,U85	0.5030579
L90,U80	0.5167806

Figura 4.4: *Se muestra la tabla con el promedio de la desviación estándar para 10 materias y 6 diferentes intervalos.*

Para la tercera prueba elegimos, de la Figura 4.4, los intervalos que tuvieran un promedio en la desviación estándar menor a 0.5. Seleccionamos otras 10 materias: *Modelos de Supervivencia y de Series de Tiempo*, *Teoría del Seguro*, *Programación Entera*, *Investigación de Operaciones*, *Geometría Moderna I*, *Geometría Analítica II*, *Lógica Matemática I*, *Cálculo*

Diferencial e Integral III, *Inferencia Estadística* y *Manejo de Datos*. La tabla con el promedio de la desviación estándar de sus datos se puede ver en la Figura 4.5.

Intervalo	Promedio_sd
L90,U85	0.4133900
L80,U90	0.4292204
L85,U80	0.4292348
L85,U90	0.4410803

Figura 4.5: *Se muestra la tabla con el promedio de la desviación estándar para 10 materias y 4 diferentes intervalos.*

Podemos ver que los valores de la Figura 4.5 son muy parecidos entre sí. Debido a ésto, hicimos otra prueba con los mismos intervalos pero con 5 materias obligatorias y con muchos alumnos. Las materias que elegimos fueron: *Geometría Analítica I*, *Cálculo Diferencial e Integral II*, *Mercados Financieros y Valuación de Instrumentos*, *Probabilidad II* y *Procesos Estocásticos I*. Hicimos la prueba para ver si había alguna diferencia en los datos y poder elegir un solo intervalo. La tabla con el promedio de la desviación estándar de sus datos se puede ver en la Figura 4.6.

Intervalo	Promedio_sd
L85,U80	0.5829679
L90,U85	0.6027183
L80,U90	0.6127408
L85,U90	0.6260881

Figura 4.6: *Se muestra la tabla con el promedio de la desviación estándar para 5 materias y 4 diferentes intervalos.*

Analizando la información de las Figuras 4.5 y 4.6, decidimos elegir los valores de $q_1 = 85$ y $q_2 = 80$. En la Figura 4.7 se muestra el intervalo formado. De dicho intervalo vamos a obtener los valores para simular la demanda de alumnos del siguiente semestre, para cada materia en cada hora.

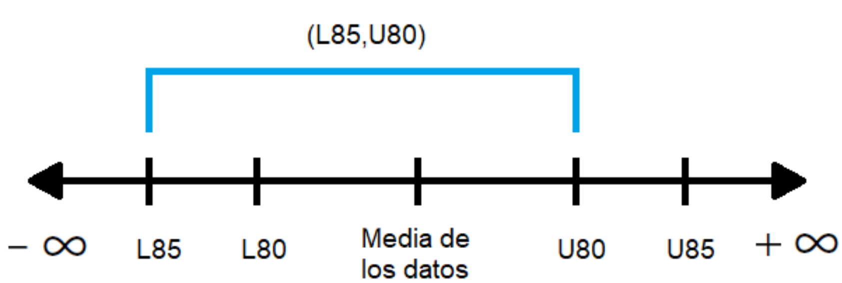


Figura 4.7: *Se muestra un diagrama con el intervalo del que se va a obtener el número de alumnos para la simulación de cada materia en cada hora.*

Finalmente, con los valores de $q_1 = 85$ y $q_2 = 80$ hicimos una prueba aleatoria (eliminando la semilla). Las materias que elegimos para dicha prueba fueron: *Modelos de Supervivencia y de Series de Tiempo, Teoría del Seguro, Cálculo Diferencial e Integral I,II y III, Investigación de Operaciones, Geometría Moderna I, Geometría Analítica II, Lógica Matemática I, Probabilidad I y II, Inferencia Estadística, Manejo de Datos, Matemáticas Financieras* y *Procesos Estocásticos I*. En la Figura 4.8 podemos ver los resultados de la prueba aleatoria mencionada. El promedio de la desviación estándar del error relativo para todas las materias es 0.48.

Materia	Intervalo	Min	Media	Max	sd
Modelos de Supervivencia y de Series de Tiempo	L85,U80	-2.4053333	-0.062628178	1.0000000	0.7625277
Teoría del Seguro	L85,U80	-0.7475410	-0.007189294	0.9685714	0.2066836
Cálculo Diferencial e Integral I	L85,U80	-1.8055556	-0.150509219	0.8544828	0.5743739
Investigación de Operaciones	L85,U80	-1.5109589	-0.113587042	0.5523364	0.4197767
Geometría Moderna I	L85,U80	-1.0729730	0.062571864	1.0000000	0.3758780
Geometría Analítica II	L85,U80	-1.6896552	-0.075029650	1.0000000	0.6112092
Lógica Matemática I	L85,U80	-1.4857143	0.009737679	1.0000000	0.4441402
Cálculo Diferencial e Integral III	L85,U80	-1.6142857	-0.118954103	0.8689189	0.5222578
Inferencia Estadística	L85,U80	-1.8022222	-0.045271946	0.9751515	0.5525782
Manejo de Datos	L85,U80	-0.5000000	0.030660747	0.8080000	0.1858422
Matemáticas Financieras	L85,U80	-0.8403974	0.030471884	0.8584270	0.2544040
Cálculo Diferencial e Integral II	L85,U80	-3.2192308	-0.053958329	1.0000000	0.6681179
Probabilidad I	L85,U80	-1.6750000	-0.047856672	0.9188034	0.4289253
Probabilidad II	L85,U80	-1.9750000	-0.124282718	1.0000000	0.7012144
Procesos Estocásticos I	L85,U80	-1.8096154	-0.109287138	0.7419643	0.5144173

Figura 4.8: *Se muestra en cada renglón la materia y el intervalo del que se tomaron los valores para la simulación. Los datos de la tabla están basados en la matriz de diferencias relativas.*

4.3. Obtención de nombres de profesores

Antes de iniciar la simulación de las solicitudes hechas (elección de materia y de horario), primero obtuvimos información de los profesores. Guardamos dicha información en la matriz *mat_nom_prof_total*, la cual tiene 2 columnas.

En la primer columna se tienen los nombres de todos los profesores que han impartido clase desde el semestre 2015-1 hasta el 2020-1. Dichos nombres los obtuvimos de la matriz *m_grande_2015*. Ésta es una submatriz de *m_grande_total* con los datos de los semestres del 2015-1 al 2020-1.

En la segunda columna de *mat_nom_prof_total* se tiene un 1 si el profesor es de tiempo completo y un 0 si es de asignatura. En las siguientes subsecciones veremos cómo llenamos esta columna y cómo hicimos la limpieza de los nombres de los profesores.

4.3.1. Profesores de tiempo completo

Para llenar la segunda columna de la matriz *mat_nom_prof_total* ingresamos a la página `www.matematicas.unam.mx/index.php/nosotros/profesores-de-tiempo-completo`, del Departamento de Matemáticas. Con la aplicación *SelectorGadget* seleccionamos uno de los nombres de los profesores de tiempo completo. En la Figura 4.9 podemos ver el código CSS que utilizamos para obtener los datos en *R*. También observamos que se seleccionaron 94 profesores.

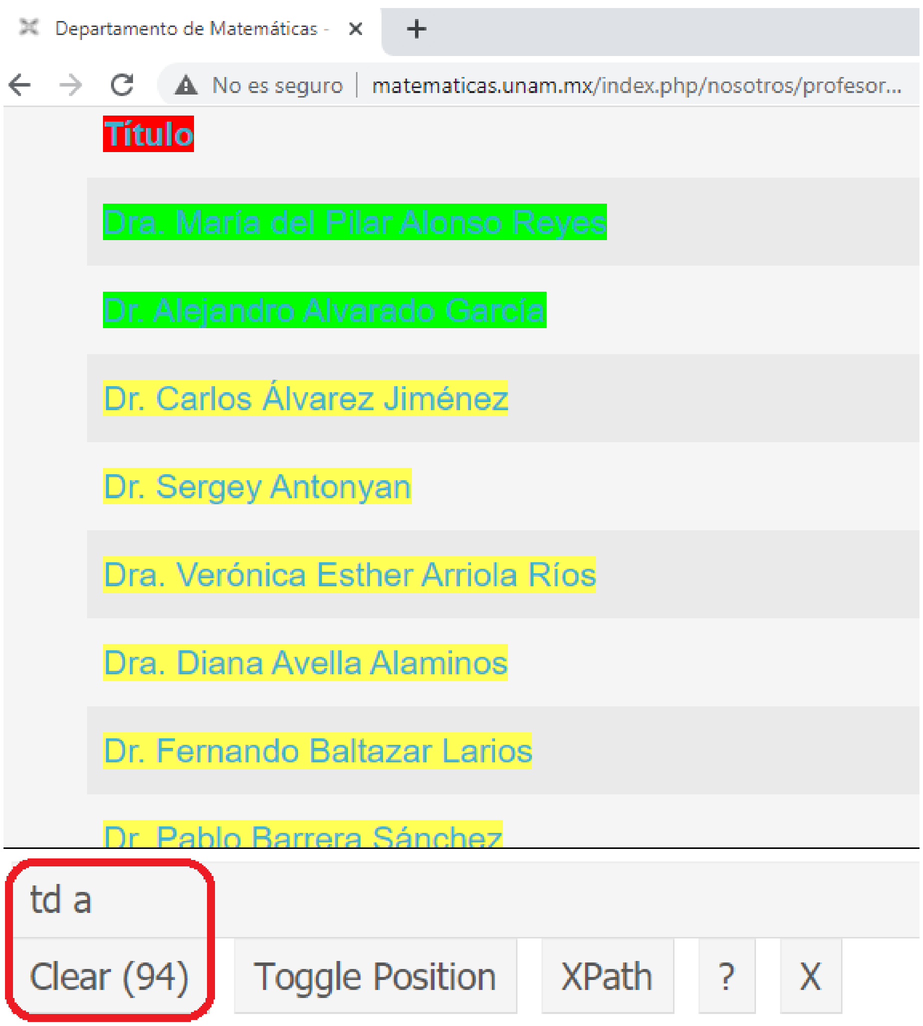


Figura 4.9: *Se muestra la selección de profesores de tiempo completo con la aplicación SelectorGadget. Se puede ver el código CSS utilizado en R.*

Al extraer la información en *R* obtuvimos un vector con 94 entradas. En la Figura 4.10 podemos ver los primeros 20 valores del vector. Notamos que cada entrada del vector inicia con los caracteres $\backslash n\backslash t\backslash t\backslash t\backslash t\backslash t\backslash t$. Estos caracteres, en la presentación final de la página de internet, indican un salto de línea y las tabulaciones o espacios que se tienen de izquierda a derecha.

```
 [1] "\n\t\t\t\t\t\t\t Dra. María del Pilar Alonso Reyes"
 [2] "\n\t\t\t\t\t\t\tDr. Alejandro Alvarado García"
 [3] "\n\t\t\t\t\t\t\t Dr. Carlos Álvarez Jiménez "
 [4] "\n\t\t\t\t\t\t\t Dr.  Sergey Antonyan"
 [5] "\n\t\t\t\t\t\t\tDra. Verónica Esther Arriola Ríos"
 [6] "\n\t\t\t\t\t\t\tDra. Diana Avella Alaminos"
 [7] "\n\t\t\t\t\t\t\tDr. Fernando Baltazar Larios"
 [8] "\n\t\t\t\t\t\t\tDr. Pablo Barrera Sánchez"
 [9] "\n\t\t\t\t\t\t\tDr. Fernando Brambila Paz"
[10] "\n\t\t\t\t\t\t\tM. en C. Alejandro Bravo Mojica"
[11] "\n\t\t\t\t\t\t\t Dra. Gabriela Campero Arena  "
[12] "\n\t\t\t\t\t\t\t Dr. Humberto Andrés Carrillo Calvet "
[13] "\n\t\t\t\t\t\t\t Dr.  Fidel Casarrubias Segura"
[14] "\n\t\t\t\t\t\t\tMat. Margarita Elvira Chávez Cano"
[15] "\n\t\t\t\t\t\t\tM. en C. Elena de Oteyza de Oteyza"
[16] "\n\t\t\t\t\t\t\tDra. Guillermina Eslava Gómez "
[17] "\n\t\t\t\t\t\t\tDra. María de Lourdes Esteva Peralta "
[18] "\n\t\t\t\t\t\t\tDr. Manuel Jesús Falconi Magaña  "
[19] "\n\t\t\t\t\t\t\tDra. Ma. Asunción Begoña Fernández Fernández"
[20] "\n\t\t\t\t\t\t\tAct.  Javier Fernández García"
```

Figura 4.10: *Se observan las primeras 20 entradas del vector obtenido con la aplicación SelectorGadget al dar click en el nombre de un profesor de tiempo completo.*

Limpiamos los datos para obtener un vector que sólo tuviera los nombres de los profesores, sin su título. Eliminamos el título porque en los horarios publicados en las páginas de la Facultad sus nombres no tienen título. También eliminamos los espacios finales que había en algunos nombres.

De esta manera obtuvimos el vector con el nombre de los profesores de tiempo completo del Departamento de Matemáticas. Dicho vector lo comparamos con la primer columna de la matriz *mat_nom_prof_total*, cuando los nombres coincidieron, pusimos un 1 en el renglón correspondiente.

Al limpiar los datos encontramos 11 nombres que analizamos de manera particular porque no aparecía el 1 en su respectivo renglón. Encontramos que no aparecía la información necesaria en la matriz *mat_nom_prof_total* por diferencias en los nombres. Encontramos diferencias por acentos, por mayúsculas y por nombres incompletos. En la Tabla 4.2 vemos los nombres que aparecen en las páginas de la Facultad comparados con los que aparecen en la página del Departamento de Matemáticas.

Nombre en páginas de la Facultad	**Nombre en página del Depto. de Matemáticas**
Alejandro Ricardo Garciadiego Dantan	Alejandro Ricardo Garciadiego Dantán
Edith Corina Sáenz Valadez	Edith Corina Sáenz Valadéz
Emilio Esteban Lluis Puebla	Emilio Lluis Puebla
Guillermo Javier Francisco Sienra Loera	Guillermo Sienra Loera
María Asunción Begoña Fernández Fernández	Ma. Asunción Begoña Fernández Fernández
María Concepción Ana Luisa Solís González-Cosío	Ana Luisa Solís González Cosío
María Isabel Puga Espinosa	Isabel Puga Espinosa
María Lourdes Velasco Arreguí	María de Lourdes Velasco Arregui
Mucuy-Kak del Carmen Guevara Aguirre	Mucuy-kak del Carmen Guevara Aguirre
Oscar Alfredo Palmas Velasco	Óscar Alfredo Palmas Velasco
Úrsula Xiomara Iturrarán Viveros	Úrsula Iturrarán Viveros

Tabla 4.2: *Se muestran los 11 nombres de los profesores de tiempo completo que se analizaron de manera individual. Se encontraron diferencias en acentos, mayúsculas y nombres incompletos.*

4.3.2. Profesores de asignatura

Al llenar la matriz *mat_nom_prof_total* con los nombres de los profesores vimos que la dimensión de dicha matriz es 1387×2. Por lo que tenemos 1387 nombres de profesores de los cuales 94 son profesores de tiempo completo. En esta subsección explicaremos cómo hicimos la limpieza de los nombres de los profesores de asignatura. Es decir los 1293 nombres que nos falta por analizar.

Lo primero que hicimos fue ordenar los nombres de los profesores de asignatura alfabéticamente. Con ellos definimos el vector *vec_prof_asig*. Al ordenarlos, encontramos 9 nombres que tenían un “ / ” al inicio de su nombre. Quitamos ese caracter y los espacios que tenía antes y después. Ordenamos nuevamente los nombres alfabéticamente. Después buscamos los nombres que tenían añadidos los nombres de los ayudantes. Dejamos únicamente el primer nombre. Aplicamos, al vector, la función `unique()` en *R*.

Con el proceso descrito obtuvimos un vector con 1246 nombres. Para comparar los nombres de los profesores, utilizamos la función `stringsim(nom_prof_1,nom_prof_2)`. Dicha función arroja el porcentaje de similitud entre los parámetros que recibe, en este caso dos nombres de profesores.

Para observar las posibles repeticiones guardamos en una matriz los nombres del vector *vec_prof_asig* y aquellos nombres con más del 60 % de coincidencia. Eliminamos 117 repeticiones de nombres. Hubo algunos casos en los que los nombres repetidos eran idénticos y en otras ocasiones diferían por acentos o por guiones. En la Tabla 4.3 vemos los nombres de los profesores que eliminamos por diferencia de acentos o guiones o nombre incompleto.

Nombre a utilizar	**Nombre eliminado**
Antonmaria Gerolamo Enrico Minzoni Alessio	Antonmaria Minzoni Alessio
Araceli Arteaga Jiménez	Aracely Arteaga Jiménez
José de Jesús Carlos Quintanar Sierra	José Jesús Carlos Quintanar Sierra
Juan Manuel Eugenio Ramírez de Arellano Niño-Rincón	Juan Manuel Eugenio Ramírez de Arellano Niño Rincón
Loiret Alejandría Dosal Trujillo	Loiret Alejandria Dosal Trujillo
María Susana Barrera Ocampo	Ma. Susana Barrera Ocampo
Manuel de Llano de la Garza	Manuel De Llano De la Garza
Mónica Alicia Clapp Jiménez-Labora	Mónica Alicia Clapp Jiménez Labora
Omar Antolín Camarena	Omar Antolin Camarena
Roberto Carrillo Lárraga	Roberto Carrillo Larraga
Rocío Jáuregui Renaud	Rocío Jauregui Renaud
Rodrigo Domínguez López	Rodrígo Domínguez López
Rosalío Fernando Rodríguez Zepeda	Rosalio Fernando Rodríguez Zepeda

Tabla 4.3: *Se muestran los nombres de los profesores de asignatura que se eliminaron por estar repetidos a causa de diferencias en el nombre como acentos, guiones o nombre incompleto.*

Finalmente obtuvimos un vector con 1129 nombres de los profesores de asignatura. Guardamos los nombres en la matriz *mat_nom_prof_total*. Dicha matriz contiene la información de 1223 profesores.

Algunas notas a considerar de esta matriz son:

- Puede haber profesores que ya no impartan clases en la Facultad.
- Puede ocurrir que no se recopile toda la información de los profesores en la Tabla 4.3 por no haber coincidencias en los nombres.

- Encontramos los nombres *Jonás Raffael Martínez Sánchez* y *Rafael Martínez Sánchez* los cuales consideramos que son nombres de personas distintas.

4.4. Simulación de tamaño de grupos

En esta sección vamos a explicar cómo hicimos la simulación del tamaño de grupos. Vamos a definir al tamaño de un grupo como el número de alumnos que va a tener cada grupo.

Con el siguiente procedimiento simulamos el tamaño de los grupos con respecto a los profesores. En la vida real cuando un alumno decide inscribirse a una materia a cierta hora, la decisión que toma para elegir el grupo al que se quiere inscribir es el profesor con el que le gustaría tomar esa materia a esa hora.

Hicimos una función en *R* que realiza los siguientes pasos:

1. Obtener, de *m_grande_2015*, la información del número de alumnos que ha tenido un profesor.
2. Tomar el mínimo (a) y el máximo (b) de esos datos.
3. Generar un número aleatorio con distribución uniforme en ese intervalo con la función `runif(1,min = a, max = b)` en *R*.
4. Redondear el número aleatorio con la función `ceiling()` en *R*.
5. Regresar el número redondeado.

Decidimos realizar de esta manera la simulación porque queremos que el número de alumnos de cada grupo dependa de los profesores y no de la distribución general que tiene el tamaño de los grupos (ver Sección 3.4).

4.5. Simulación de solicitudes de profesores

En esta sección vamos a explicar cómo hicimos la simulación de la solicitud de los profesores. En la vida real los profesores pueden elegir libremente las materias que quieren impartir y seleccionan las horas a las que desean impartir sus clases. Dado que no contamos con esa información decidimos simular la elección de materias y horarios en base a la información que tenemos de semestres anteriores.

Como vimos en la Figura 4.1 simulamos dos veces las solicitudes de los profesores, en el proceso de asignación. A la primera vez que simulamos las solicitudes la llamaremos *Solicitud oculta* y a la segunda la llamaremos *Solicitud pseudo-real*. La explicación de su uso lo vemos a continuación.

- Solicitud oculta: La llamamos oculta porque nos ayuda para la generación de los esqueletos. No influye directamente en la asignación final.
- Solicitud pseudo-real: Es la simulación de las posibles elecciones que los profesores harían en la vida real. Nos ayuda directamente a realizar la asignación final.

El procedimiento para ambos casos es prácticamente el mismo. Primero obtuvimos una matriz, llamada *mat_1_solicitud*, la cual tiene la información de la solicitud de un profesor. La matriz tiene 5 columnas *(Profesor, TC, Materia, Num_Materia, Horario)* y 6 renglones. Los pasos que realizamos para obtener la matriz *mat_1_solicitud*, con la solicitud de un profesor, son los siguientes:

1. Llenar la columna *Profesor* con el nombre del profesor del cual queremos realizar la solicitud.

2. Llenar la columna *TC* dependiendo del tipo de profesor que se haya elegido en el paso anterior. Esta columna tiene un 1 en cada renglón si el profesor es de tiempo completo y un 0 si el profesor es de asignatura.

3. Obtener, de *m_grande_2015*, la información de las materias que ha impartido el profesor elegido. Guardar la información en el vector *materias_profesor*. Se tienen 3 casos con respecto al número de materias que tiene el vector:

    a) El número de materias es 2: Llenar los primeros 3 renglones, de la columna *Materia*, con la información de la materia 1 y los últimos 3 renglones con la información de la materia 2.

    b) El número de materias es mayor o igual a 3: Se toma una muestra de dos materias, con la función `sample(materias_profesor, size = 2)` en *R*. Se llena la columna *Materia* como el caso anterior.

    c) El número de materias es 1: Llenar la columna *Materia* con esa materia.

4. Llenar la columna *Num_Materia* de *mat_1_solicitud* con los números de materia correspondientes a las elegidas en el paso anterior.

5. Obtener, de *m_grande_2015*, la información de las horas en las que ha dado clases el profesor elegido. Guardar la información en el vector *horas_profesor*. Se tienen 4 casos con respecto al número de horas que se encuentran en el vector:

    a) El número de horas es 3: Llenar los renglones 1 y 4, de la columna *Horario*, con la información de la hora 1. Llenar los renglones 2 y 5 con la información de la hora 2. Llenar los renglones 3 y 6 con la información de la hora 3.

    b) El número de horas es mayor o igual a 4: Se toma una muestra de 3 horas, con la función `sample(horas_profesor, size = 3)` en *R*. Se llena la columna *Horario* como el caso anterior.

    c) El número de horas es 2: Llenar los renglones 1,2,4 y 5, de la columna *Horario*, con la información de la hora 1. Llenar los renglones 3 y 6 con la información de la hora 2.

    d) El número de horas es 1: Llenar la columna *Horario* con esa hora.

En la Figura 4.11 vemos un ejemplo de la matriz *mat_1_solicitud*.

	Profesor	TC	Materia	Num_Materia	Horario
1	Margarita Elvira Chávez Cano	1	Modelos no Paramétricos y de Regresión	57	9
2	Margarita Elvira Chávez Cano	1	Modelos no Paramétricos y de Regresión	57	10
3	Margarita Elvira Chávez Cano	1	Modelos no Paramétricos y de Regresión	57	11
4	Margarita Elvira Chávez Cano	1	Modelos de Supervivencia y de Series de Tiempo	123	9
5	Margarita Elvira Chávez Cano	1	Modelos de Supervivencia y de Series de Tiempo	123	10
6	Margarita Elvira Chávez Cano	1	Modelos de Supervivencia y de Series de Tiempo	123	11

Figura 4.11: *Se observa un ejemplo de la matriz mat_1_solicitud para un profesor de tiempo completo.*

El proceso se repite para cada uno de los profesores en la matriz *mat_nom_prof_total* obtenida en la Sección 4.3. La matriz formada con las solicitudes de todos los profesores la llamamos *mat_solicitudes*. A ella le quitamos los renglones repetidos. Con estos pasos realizamos la *solicitud oculta*. Ésta nos sirve para poder simular adecuadamente los esqueletos.

Para realizar la *solicitud pseudo-real* hacemos una comparación del esqueleto generado y la matriz *mat_solicitudes*. Cabe señalar que la *solicitud pseudo-real* depende del esqueleto simulado y la *solicitud oculta* no.

4.6. Simulación de la demanda de alumnos

La demanda del número de alumnos para el siguiente semestre la hicimos por materia y por hora. Para poder realizar la simulación lo primero que hicimos fue acomodar la información que teníamos por semestres y por hora. El procedimiento que seguimos fue el siguiente:

1. Definir el semestre del cual se quiere obtener la simulación (*sem_sig*).
2. Definir el número de semestres que se quieren como ventana de información (k).
3. Tomar una submatriz de *m_grande_total* con la información de una materia para los semestres en la ventana de información.
4. Para cada semestre dentro de la ventana de información se suma el número de alumnos en cada hora.
5. Se obtiene una matriz de $t \times k$ como la que se puede ver en la Figura 4.12. Recordemos que $t = 15$ y representa el número de horas en las que se imparten clases.

Con el procedimiento descrito pudimos generar vectores por hora. Aplicamos la función `hw()` en *R* para obtener la demanda de alumnos esperados para el siguiente semestre. En la Figura 4.13 vemos la matriz vista en la Figura 4.12 junto con el vector de alumnos simulados (señalado en rojo). El vector contiene la demanda de alumnos simulados para el semestre 2020-2 de la materia *Modelos de Supervivencia y Series de Tiempo*.

Notamos que el valor de la demanda de alumnos es cero cuando en todos los semestres de alguna hora no hay datos. En el ejemplo, es el caso de las 7hrs, 8hrs, 13hrs, 14hrs, 15hrs, 16hrs y 21hrs.

	20181	20182	20191	20192	20201
7-8	0	0	0	0	0
8-9	0	0	0	0	0
9-10	0	0	71	0	52
10-11	198	0	75	0	144
11-12	0	44	0	9	0
12-13	0	75	0	97	0
13-14	0	0	0	0	0
14-15	0	0	0	0	0
15-16	0	0	0	0	0
16-17	0	0	0	0	0
17-18	0	52	0	40	47
18-19	0	0	0	0	88
19-20	78	0	63	0	0
20-21	0	53	79	69	0
21-22	0	0	0	0	0

Figura 4.12: *Se puede ver la información del número de alumnos reales de la materia "Modelos de Supervivencia y Series de Tiempo" por semestre y por hora.*

Observando los datos de las 10hrs. vemos que en los semestres pares no hay alumnos, por lo que en la simulación se obtiene únicamente un alumno. Si vemos los datos de las 17hrs vemos que de los 5 semestres en la ventana se tienen alumnos en los semestres pares y en un semestre impar, el número de alumnos simulados para esa hora son 31 alumnos.

Con estos ejemplos podemos ver de manera tangible que el modelo respeta la estacionalidad semestral que tienen los datos.

	20181	20182	20191	20192	20201	20202
7-8	0	0	0	0	0	0
8-9	0	0	0	0	0	0
9-10	0	0	71	0	52	61
10-11	198	0	75	0	144	1
11-12	0	44	0	9	0	8
12-13	0	75	0	97	0	122
13-14	0	0	0	0	0	0
14-15	0	0	0	0	0	0
15-16	0	0	0	0	0	0
16-17	0	0	0	0	0	0
17-18	0	52	0	40	47	31
18-19	0	0	0	0	88	29
19-20	78	0	63	0	0	6
20-21	0	53	79	69	0	132
21-22	0	0	0	0	0	0

Figura 4.13: *En esta figura se señala en rojo el vector con la demanda simulada para el 2020-2 de la materia "Modelos de Supervivencia y Series de Tiempo".*

Obtuvimos vectores con la demanda simulada para cada una de las materias y formamos una matriz de $t \times m$, llamada *mat_demanda_alumnos*. Recordemos que m es el número de materias que se van a impartir. En la Figura 4.14 podemos ver un ejemplo de cómo se ve la matriz formada.

Analicemos 2 pares de grupos, primero veamos la columna de *Álgebra Superior II* (2) y la de *Geometría Analítica I* (4). Ambas son materias obligatorias para Actuaría, Matemáticas y Matemáticas Aplicadas. La primera corresponde a semestres pares y la segunda a semestres impares. Notamos que para *Geometría Analítica I*, se tienen alumnos prácticamente en cada hora, pero el número no es muy grande. Para *Álgebra Superior II* hay varias horas con cero alumnos simulados pero hay dos grandes cantidades, una a las 9hrs con 832 alumnos y la otra a las 18hrs con 224 alumnos. Con esta comparación podemos ejemplificar la diferencia entre una materia que corresponde a semestres pares y una de semestres impares.

Ahora analicemos las columnas de *Seminario de Topología A* (3) y *Probabilidad II* (6). La primera es una materia optativa para Matemáticas. La segunda es una materia obligatoria para Actuaría, correspondiente a semestres pares y optativa para Ciencias de la Computación, Matemáticas y Matemáticas Aplicadas. El número total de alumnos simulados para *Seminario de Topología A* es menor a 20, en cambio para *Probabilidad II* se tiene una gran cantidad de alumnos a las 8hrs, 9hrs y 10hrs. Considerando los valores que se tienen en el turno vespertino para *Probabilidad II*, notamos que a las 19hrs también hay una gran cantidad de alumnos. Con esta comparación podemos ejemplificar la diferencia entre una materia obligatoria y una optativa, así como la diferencia entre el turno matutino y vespertino.

	Topología I	Álgebra Superior II	Seminario de Topología A	Geometría Analítica I	Geometría Moderna I	Probabilidad II
7-8	0	16	0	37	0	0
8-9	2	59	0	39	53	264
9-10	0	832	0	38	55	160
10-11	15	0	1	16	0	187
11-12	56	0	2	0	0	5
12-13	0	0	1	12	6	0
13-14	2	0	5	30	19	0
14-15	32	8	0	107	86	0
15-16	20	7	0	43	1	0
16-17	0	0	0	0	36	0
17-18	1	0	10	27	0	8
18-19	44	224	0	187	0	9
19-20	0	0	0	45	0	85
20-21	0	9	0	10	16	0
21-22	0	6	0	16	0	0

Figura 4.14: *Se muestra una submatriz de "mat_demanda_alumnos". Los vectores contienen el número de alumnos simulados para el semestre 2020-2 por hora de algunas materias.*

4.7. Modelo de Mezcla Gaussiana

El modelo de Mezcla Gaussiana también es llamado modelo de mezcla de normales. En él se tienen dos o más distribuciones normales. Se hace una combinación lineal de ellas y se obtiene una nueva distribución de probabilidad.

En *R* se puede obtener dicha distribución con la función `normalmixEM()`. La función regresa un objeto de tipo *mixEM* el cual contiene el modelo de una mezcla de distribuciones normales obtenidas por el algoritmo de maximización de la esperanza (EM - *expectation maximization algorithm*).

Do Chuong y Batzoglou nos indican, en su artículo *What is the expectation maximization algorithm?* [4], que el algoritmo EM es una generalización natural de la estimación por máxima verosimilitud. Ésto para el caso en donde se tiene información incompleta. Recordemos que la máxima verosimilitud se utiliza para encontrar la mejor manera de ajustar una distribución a los datos.

En el algoritmo EM, los parámetros iniciales se toman de los datos reales. Con ellos se obtienen unos parámetros finales que se convierten en los parámetros para la siguiente iteración. Así sucesivamente.

Cabe mencionar que el objeto de tipo *mixEM* contiene los valores de μ y σ finales. Éstos nos sirven para simular números aleatorios con una distribución normal. Dicha distribución tiene k medias, así como k desviaciones estándar. Los principales parámetros que recibe la función `normalmixEM()` son:

- *x:* Vector con los datos a los que se les quiere aplicar el modelo.
- *mu:* Vector con las medias iniciales para el algoritmo EM.
- *sigma:* Vector con las desviaciones estándar iniciales del modelo.
- *k:* Número de distribuciones normales que se ajustan a los datos.

En la Figura 4.15 se puede ver el histograma con el número de alumnos esperados por hora. Los datos corresponden a una matriz *mat_demanda_alumnos* (ver Sección 4.6). Con el comando `mixmdl_1_D <- normalmixEM(wait_alumnos,k = 4)` guardamos el modelo inicial. La línea azul de dicha figura corresponde a la densidad ajustada de 1000 números aleatorios con una distribución normal con 4 medias. El comando para obtener la densidad es: `density(rnorm(1000,mean = mixmdl_1_D$mu,sd = mixmdl_1_D$sigma)`. Se tomaron los valores de μ y σ arrojados por el modelo *mixmdl_1_D*.

En la Figura 4.16 se puede observar el histograma con el número de alumnos esperados por hora. Los datos corresponden a 5 matrices *mat_demanda_alumnos* (ver Sección 4.6). Con el comando `mixmdl_D <- normalmixEM(wait_alumnos_final,mixmdl_1_D$mu)` guardamos el modelo final. En este caso la función recibió los datos del modelo inicial. Los valores de μ son: $7.49, 10.30, 13.78, 17.23$. Para obtener la densidad ajustada, mostrada en la figura (línea azul), utilizamos los valores de μ y σ arrojados por el modelo *mixmdl_D*.

Notamos que la densidad se ajusta mejor a los datos. Ésto debido a que para este caso se tienen mejores valores iniciales. Se puede ver el pico de las 10hrs, también se observa que se toman en cuenta los picos de las 14hrs, 16hrs y 18hrs.

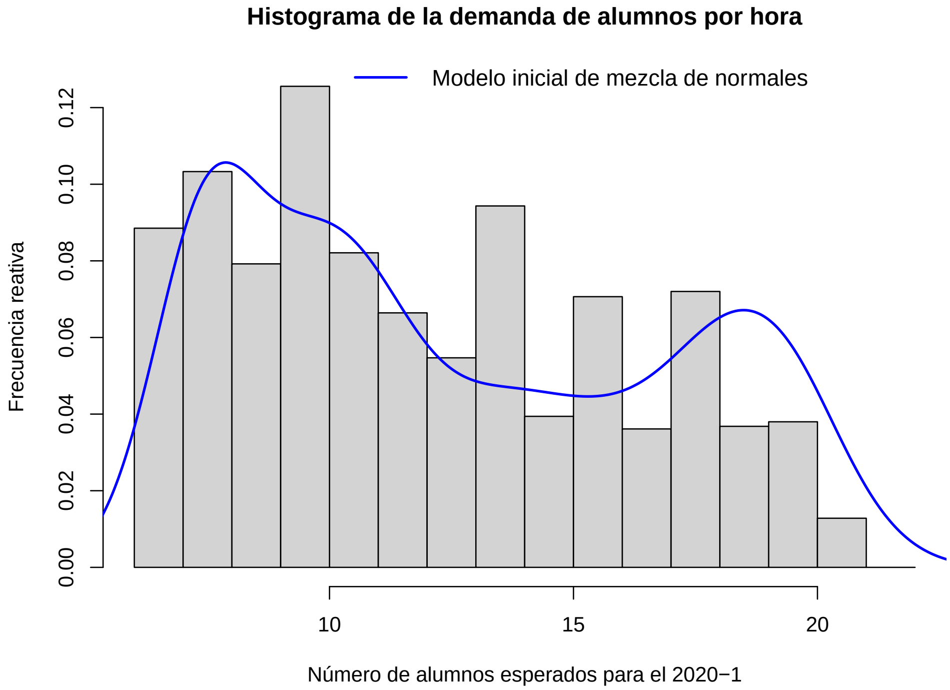


Figura 4.15: *Se muestra el histograma del número de alumnos esperados por hora. La línea azul corresponde a la densidad ajustada de 1000 números aleatorios con la distribución obtenida con el modelo inicial de mezcla de normales.*

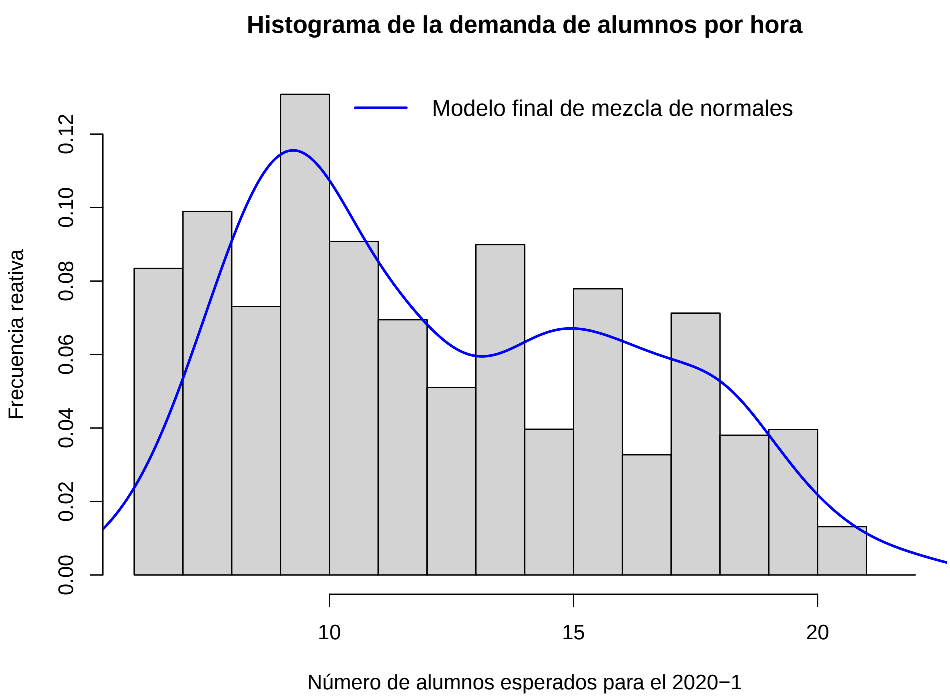


Figura 4.16: *Se muestra el histograma del número de alumnos esperados por hora. La línea azul corresponde a la densidad ajustada de 1000 números aleatorios con la distribución obtenida con el modelo final de mezcla de normales.*

El modelo de mezcla de normales lo utilizamos para simular los esqueletos de los horarios. En las siguientes secciones veremos cómo aplicamos la función `normalmixEM()` para obtener la matriz *mat_esqueleto*.

4.8. Obtención de D' y D_0

Los esqueletos que vamos a simular dependen de la demanda de alumnos. En esta sección vamos a mostrar 4 diferentes metodologías que probamos para poder simular adecuadamente los esqueletos. Ésto basándonos en el número de alumnos simulados para el siguiente semestre.

Definimos las siguientes matrices:

- D' : Matriz de $t \times m$, con la demanda simulada por alguna de las 4 metodologías.

- D^0 : Matriz de $t \times m$, con la cual se va a comparar D' para calificarla. Esta matriz se obtiene haciendo el promedio entre una matriz *mat_demanda_alumnos* (ver Sección 4.6) y una matriz de demanda de alumnos, obtenida con el modelo de mezcla de normales (ver Sección 4.7).

La calificación de las metodologías depende de la diferencia relativa entre D^0 y D'. Los pasos que seguimos para obtener las calificiones son:

1. Definir la matriz C, de $t \times m$. Esta matriz va a guardar las calificaciones por grupo de D'.

2. Para cada $C_{h,j}$ guardar el valor de $\dfrac{D^0_{h,j} - D'_{h,j}}{D^0_{h,j}}$.

3. Si $D^0_{h,j} = 0$ entonces $C_{h,j} = 1$ si faltan alumnos y $C_{h,j} = -1$ si sobran alumnos, es decir:
$$C_{h,j} = \begin{cases} 1 & \text{si } D^0_{h,j} > D'_{h,j} \\ -1 & \text{si } D^0_{h,j} < D'_{h,j} \\ 0 & \text{e.o.c.} \end{cases}$$

4. Definir el vector *vec_calif_x_materia* con el promedio por columna de C. Este vector guarda las calificaciones por materia de D'.

Para cada metodología, realizamos 10 simulaciones y calificamos las matrices D' generadas. Con este procedimiento obtuvimos 4 matrices de 10 renglones y m columnas. Graficamos cada matriz con la función `matplot()` en R. Cada gráfica contiene m líneas con 10 puntos cada línea. A continuación mostramos las 4 gráficas.

En la Figura 4.17 vemos las calificaciones por materia de la metodología A. Notamos que se encuentran entre -5 y 1. Ésto quiere decir que en promedio con este método sobra hasta un 500 % de alumnos y falta casi un 100 % al hacer la simulación.

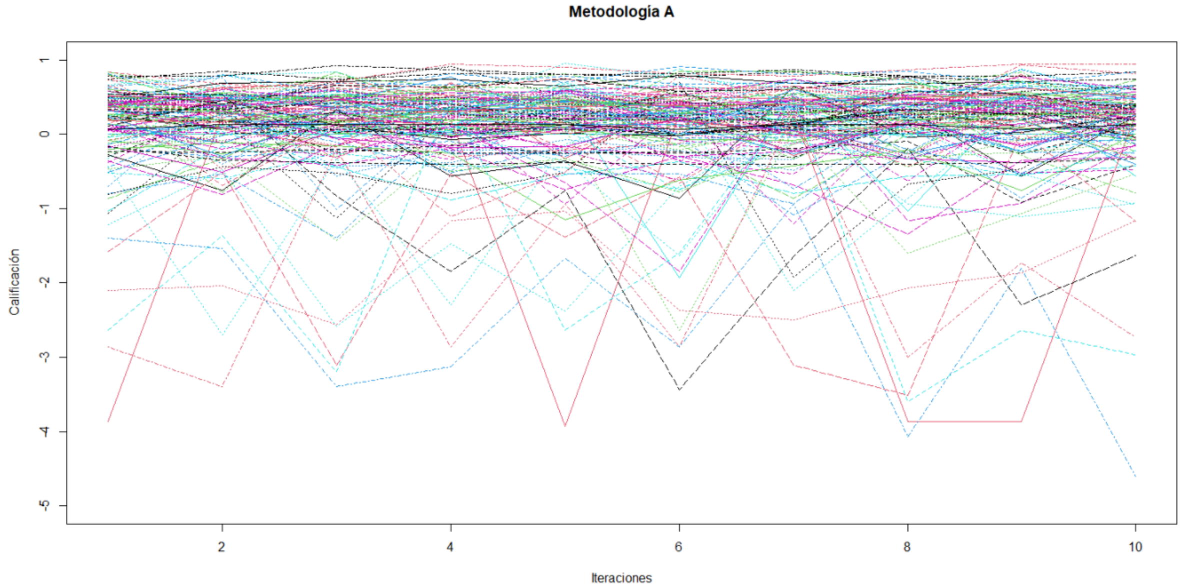


Figura 4.17: *Se muestran las calificaciones por materia de la metodología A.*

En la Figura 4.18 vemos las calificaciones por materia de la metodología B. Notamos que se encuentran entre -0.5 y 0.8. Ésto quiere decir que en promedio con este método sobra hasta un 50 % de alumnos y falta casi un 80 % al hacer la simulación.

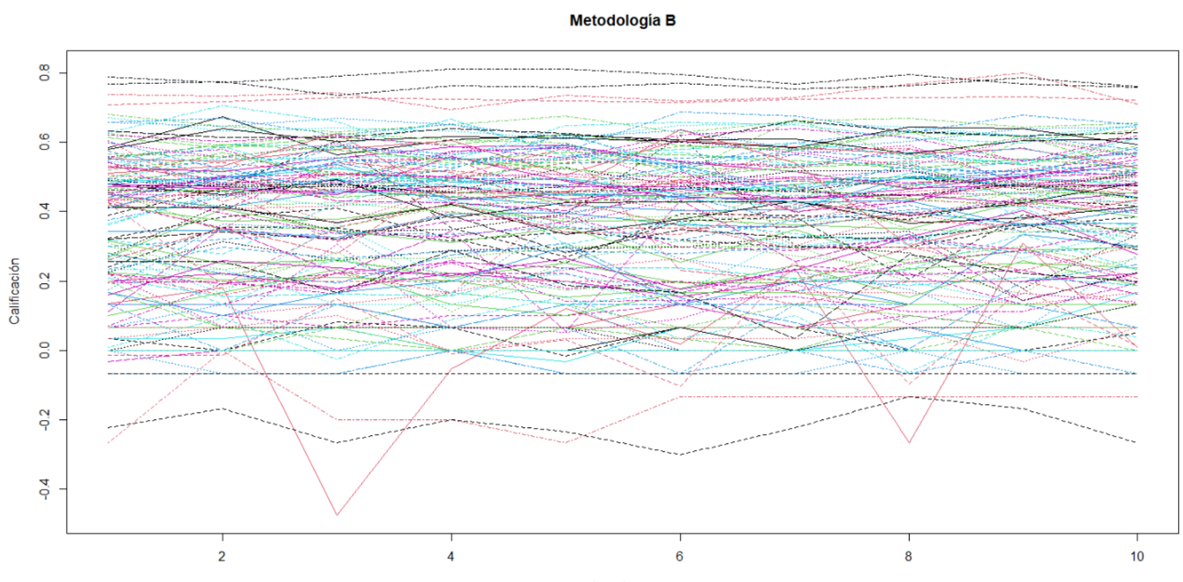


Figura 4.18: *Se muestran las calificaciones por materia de la metodología B.*

En la Figura 4.19 vemos las calificaciones por materia de la metodología C. Notamos que, al igual que en la metodología B, las calificaciones se encuentran entre -0.5 y 0.8. En este caso observamos que hay una mayor concentración de materias (líneas) entre 0.5 y 0.8. Ésto comparado con el método B que tiene una mayor concentración entre 0.4 y 0.6.

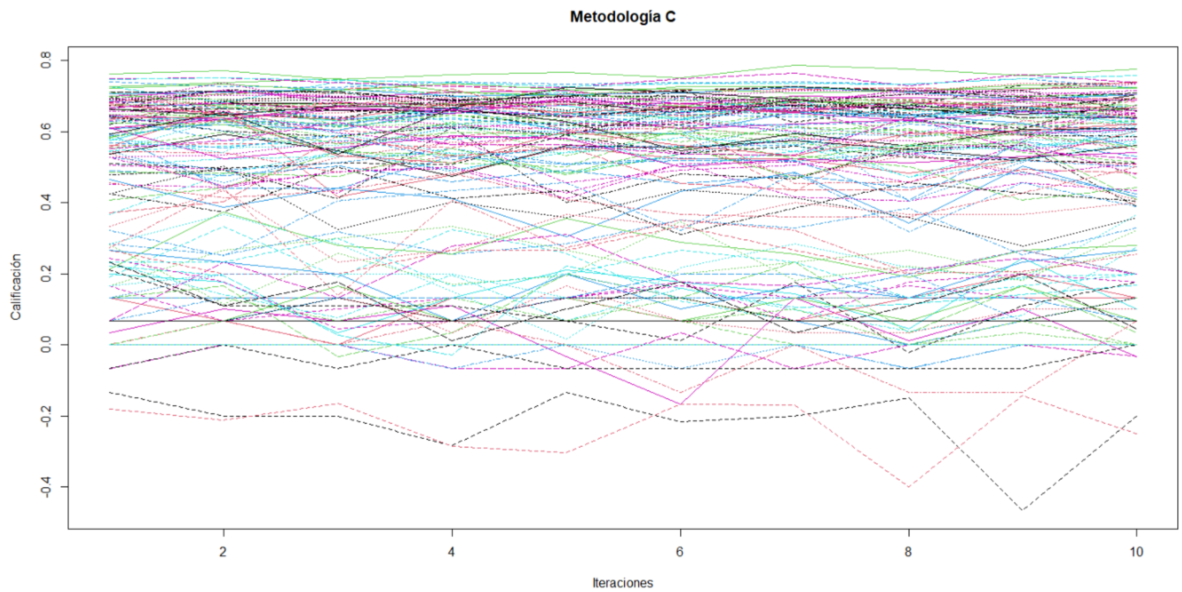


Figura 4.19: *Se muestran las calificaciones por materia de la metodología C.*

En la Figura 4.20 vemos las calificaciones por materia de la metodología D. Notamos que se encuentran entre -6 y 0.4. Ésto quiere decir que en promedio con este método sobra hasta un 600 % de alumnos y falta casi un 40 % al hacer la simulación. Podemos observar que sólo una materia tiene calificaciones por debajo de -3. En general todas se concentran entre -2.5 y 0.4.

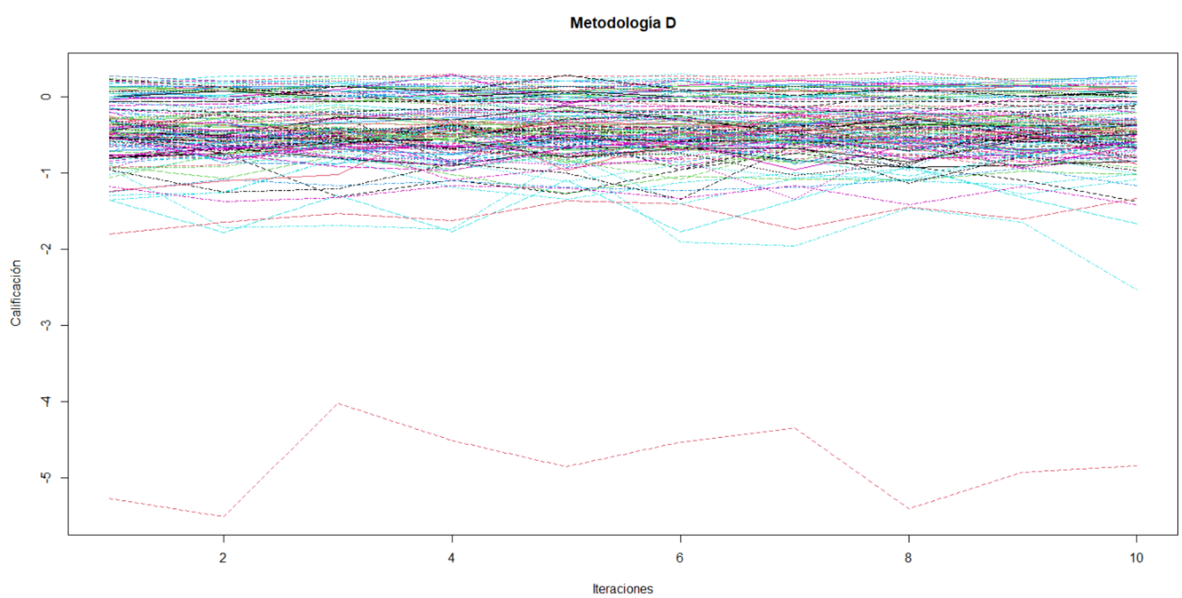


Figura 4.20: *Se muestran las calificaciones por materia de la metodología D.*

Decidimos analizar las metodologías *B* y *C* ya que son las que muestran las mejores calificaciones. Para ello graficamos las matrices de calificaciones con la función `heatmap()` en *R*. Cabe aclarar que las matrices de calificaciones están ordenadas de menor a mayor por renglones. En la Figura 4.21 vemos el correspondiente a la metodología *B*. En la Figura 4.22 vemos el *heatmap* referente a la metodología *C*.

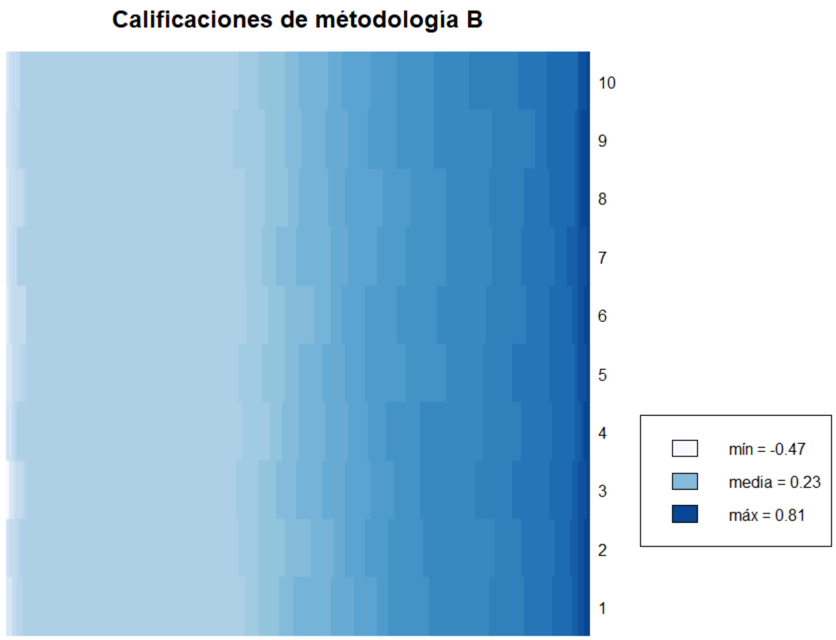


Figura 4.21: *Se muestra el heatmap de las calificaciones por materia de la metodología B.*

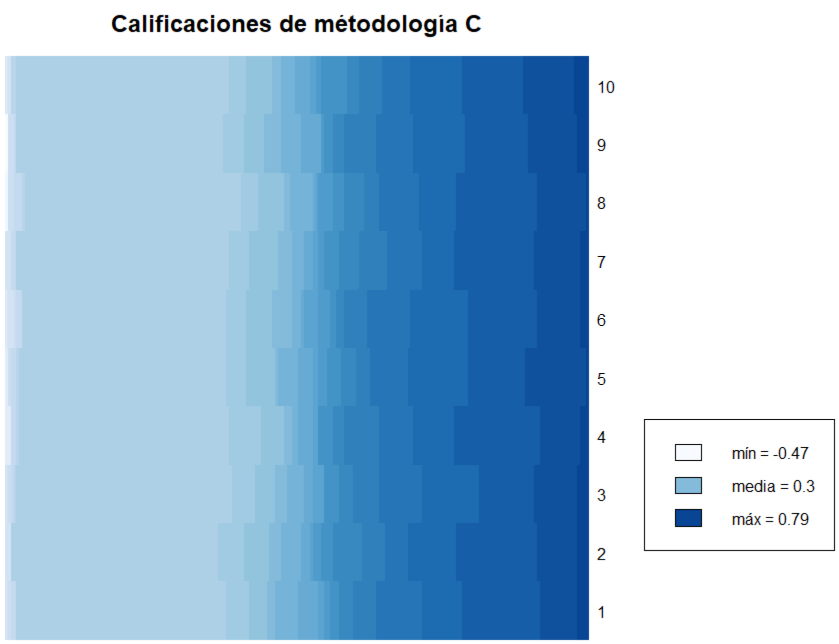


Figura 4.22: *Se muestra el heatmap de las calificaciones por materia de la metodología C.*

Para elegir entre las dos metodologías, tomamos en cuenta que el error relativo estuviera más cercano a cero. Al ver ambos *heatmaps* observamos que el correspondiente al método B es más claro que el del método C. Por lo que elegimos la metodología B para simular los esqueletos.

En la metodología B la matriz D' es una matriz *mat_demanda_alumnos*. Es por ello que se genera con el procedimiento descrito en la Sección 4.6. La matriz *mat_esqueleto* se genera con el modelo de mezcla de Normales visto en la Sección 4.7. En las siguientes secciones se describe a detalle el procedimiento que seguimos para obtener la matriz *mat_esqueleto*.

4.9. Simulación de esqueletos

Considerando que ya se generaron las matrices D' (ver Sección 4.6) y *mat_solicitudes* (ver Sección 4.5), el proceso que seguimos para obtener un esqueleto es el siguiente:

1. Elegir un profesor de tiempo completo al azar.
2. Elegir al azar un horario y una materia que haya solicitado el profesor elegido en el paso anterior. Con estos datos obtenemos las coordenadas (h, j) para las matrices D' y *mat_esqueleto*.
3. Verificar que a esa materia en esa hora aún le sobran alumnos, en la entrada (h, j) de D'.
4. Simular el número de alumnos para ese grupo (ver Sección 4.4).
5. Restar el número de alumnos simulados en el paso anterior, de la materia y hora elegidas, en la entrada (h, j) de D'.
6. Ese profesor ya no puede impartir clases a esa hora. Retirar renglones correspondientes de *mat_solicitudes*.
7. Repetir los pasos de 1 a 6 hasta que se terminen los profesores.
8. Una vez que se terminen los profesores de tiempo completo, hacer los pasos de 1 a 7 con los profesores de asignatura.

Algunas notas a considerar del procedimiento son:

- Los profesores de tiempo completo deben cumplir con sus horas, por contrato.
- Los profesores sólo pueden tener asignadas a lo más 2 materias.
- Las condiciones de paro del proceso son:
  a) Ya se cubrió toda la demanda
  b) Ya no hay más profesores
  c) Llegar a una cota predefinida para que el ciclo no se haga infinito o tarde mucho en cumplir las condiciones anteriores.

4.10. Obtención de mat_esqueletos

En esta sección vamos a explicar cómo generamos la matriz *mat_esqueleto*, utilizando la metodología B seleccionada en la Sección 4.8. La matriz tiene t renglones y m columnas. En la entrada (h, j) tiene el número de grupos simulados para la hora h y la materia j. La matriz *mat_esqueleto* depende de la demanda de alumnos y de las solicitudes de los profesores.

Los pasos que seguimos son:

1. Definir *n_rep*, el número de veces que se va a generar la matriz D'_n con la demanda de alumnos para el siguiente semestre.

2. Simular D'_1 con la función *gen_mat_demanda_alumnos*. Los pasos de esta función se describen en la Sección 4.6.

3. Definir la matriz *prom_D* igual a D'_1. La matriz *prom_D* guardará el promedio del número de alumnos simulados.

4. Simular *mat_solicitudes* con la función *gen_solicitudes*. El procedimiento de esta función está descrito en la Sección 4.5.

5. Simular un esqueleto inicial con la función *gen_esqueleto*. Los pasos de esta función se pueden ver en la Sección 4.9.

6. Guardar el número de grupos por materia.

7. Convertir y guardar los datos del esqueleto inicial para obtener la distribución por horas. Los datos se guardan en el vector *wait_mat_esqueleto*.

8. Graficar los datos del esqueleto inicial para ver su distribución. Con esta gráfica encontrar el número de medias inicial, en nuestro caso $k = 4$ (ver Figura 4.23).

9. Definir el modelo inicial *mixmdl_1_esqueleto* con el siguiente comando en *R*:

   `normalmixEM(wait_mat_esqueleto,k = 4).`

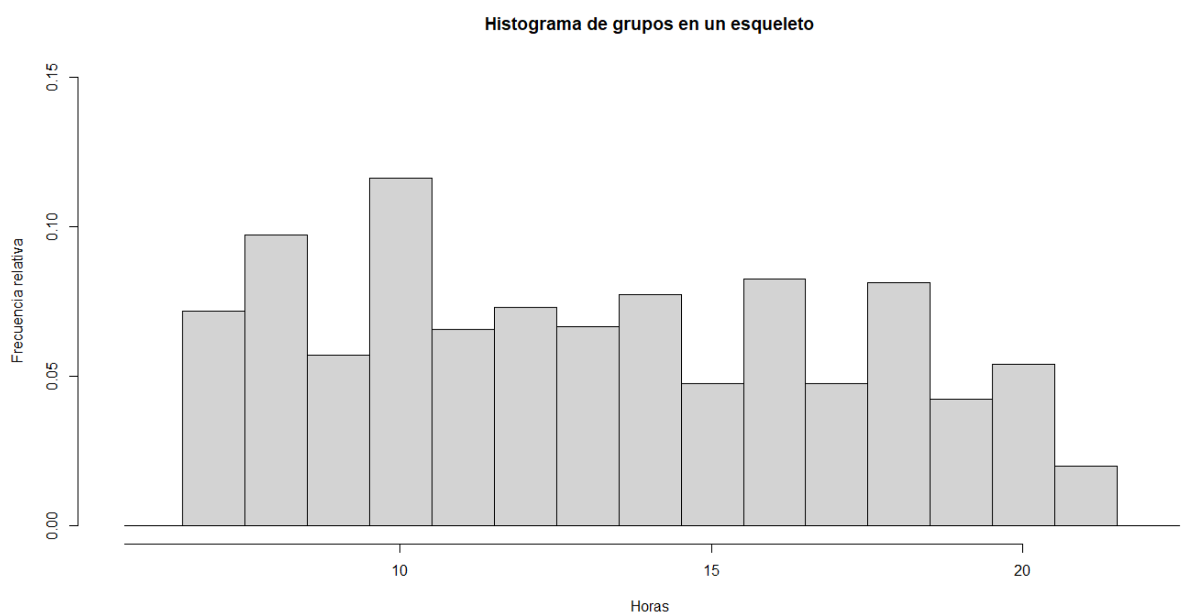


Figura 4.23: *Se muestra el histograma con el número de grupos por hora en el esqueleto inicial.*

Pasos a repetir $(n = 2, \ldots, n\_rep)$:

1. Obtener D'_n con la función *gen_mat_demanda_alumnos*.

2. Definir $prom\_D = prom\_D + D'_n$.

3. Simular *mat_solicitudes* con la función *gen_solicitudes*.

4. Simular un esqueleto con la función *gen_esqueleto*.

5. Guardar el número de grupos por materia.

6. Convertir y guardar los datos del esqueleto en el vector *wait_mat_esqueleto*.

Pasos finales:

1. Calcular el promedio de grupos por materia. Para ello, aplicar las siguientes funciones de *R*, a la matriz *prom_D*: `ceiling(colMeans((prom_D))`

2. Definir el modelo final *mixmdl_esqueleto* con el siguiente comando en *R*:

   `normalmixEM(wait_mat_esqueleto,k = 4,mean=mixmdl_1_esqueleto$mu).`

   En la Figura 4.24, se puede ver el histograma con todos los datos de los esqueletos simulados. La línea azul representa la distribución ajustada por el modelo final *mixmdl_esqueleto*.

3. Generar la matriz *mat_esqueleto* en base al promedio obtenido y a la distribución del modelo final. Por ejemplo, si se tiene una materia con 5 grupos simulados, entonces se simulan 5 números aleatorios con distribución Normal. El comando en *R* es: `round(rnorm(5,mixmdl_esqueleto$mu,mixmdl_esqueleto$sigma))`

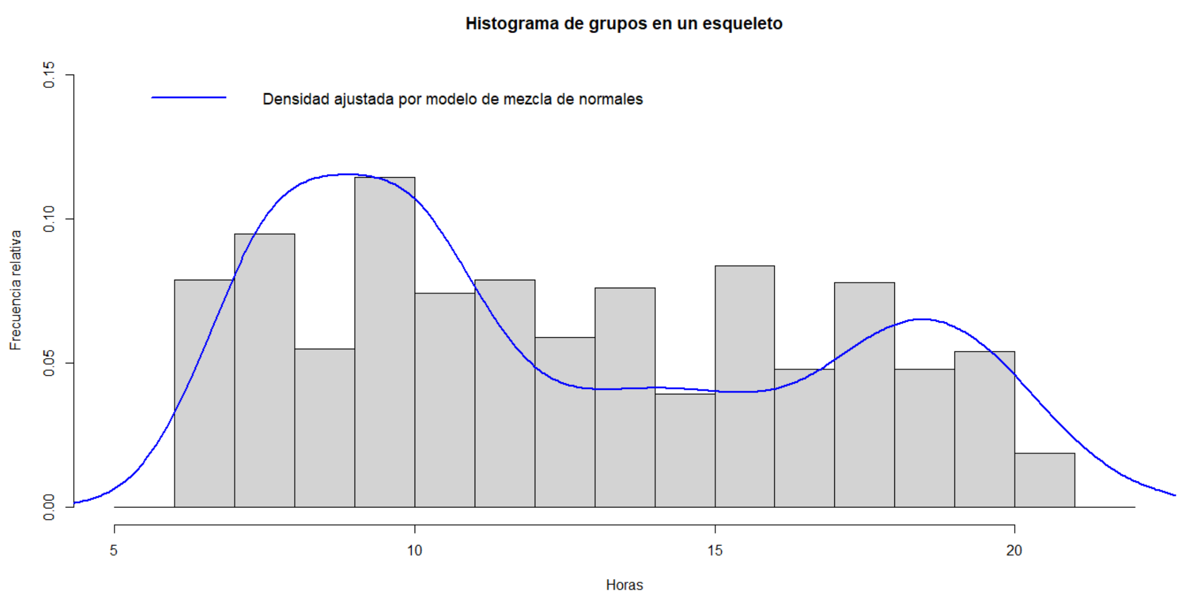


Figura 4.24: *Se muestra el histograma con el número de grupos por hora de todos los esqueletos simulados. La línea azul representa la distribución ajustada por el modelo final.*

En la Figura 4.25 vemos un ejemplo de una submatriz de *mat_esqueleto* para el semestre 2020-2. Observemos las últimas 4 columnas que corresponden a las materias de *Cálculo Diferencial e Integral I, II, III* y *IV*. Notamos que el número de grupos simulados para *Cálculo Diferencial e Integral II* es mayor a al número de grupos de *Cálculo Diferencial e Integral I*. Ésto se debe al comportamiento descrito en la Sección 3.2 y el semestre 2020-2 es par. Para *Cálculo Diferencial e Integral III* y *IV* el número de grupos simulados es prácticamente igual.

	Inferencia Estadística	Investigación de Operaciones	Teoría de Redes	Cálculo de las Variaciones	Cálculo Diferencial e Integral IV	Cálculo Diferencial e Integral I	Cálculo Diferencial e Integral II	Cálculo Diferencial e Integral III
7-8	1	1	1	1	4	2	3	3
8-9	2	3	0	0	1	2	4	2
9-10	3	0	0	0	1	1	3	1
10-11	0	2	0	0	2	2	4	3
11-12	0	0	1	0	2	2	1	3
12-13	1	1	1	0	2	0	3	1
13-14	0	1	0	0	2	0	0	1
14-15	0	0	0	0	1	1	1	0
15-16	1	0	0	0	0	1	2	1
16-17	0	1	0	0	0	0	2	1
17-18	0	0	0	0	1	0	2	0
18-19	1	1	0	0	1	0	0	3
19-20	0	1	1	0	0	0	0	0
20-21	0	0	0	0	0	2	2	0
21-22	1	0	0	0	1	0	0	0

Figura 4.25: *Se muestra una submatriz de "mat_esqueleto" para el semestre 2020-2. En la entrada (h,j) se observa el número de alumnos simulados para la hora h y la materia j.*

Capítulo 5

Algoritmo Genético

El Algoritmo Genético (AG) es un método de computación evolutiva o *machine learning*, basado en la teoría sintética de la evolución. Dicha teoría, a grandes rasgos, combina el mecanismo de la selección natural de Darwin con la genética de Mendel. Nos indica que el individuo más apto sobrevive, por lo que entre mejores sean los padres, mejor es la descendencia.

Actualmente el AG se utiliza para resolver problemas de búsqueda y optimización. Las áreas de aplicación son por ejemplo: economía, finanzas, medicina, ciencias sociales, investigación de operaciones, hidráulica, aeronáutica y química. Algunos ejemplos dentro de estas áreas son: diseño de redes de agua potable, optimización de portafolios de inversión, el problema del agente viajero y asignar asientos en un evento.

Los pasos del AG son los siguientes:

1. Selección: Se define una población de tamaño n. El valor de aptitud o adaptabilidad, de cada elemento en la población, se asigna al evaluar su utilidad en la función objetivo. Entre mejor sea el elemento, más alto será su valor de adaptabilidad. Se eligen 2 elementos de la población, llamados padres. La selección de los padres depende de qué tan aptos sean. Con dichos padres se va a formar un hijo.

2. Cruce: Con cierta probabilidad se toma información de los padres. Dicha información la llamaremos genes.

3. Mutación: Cada gen agregado al hijo tiene una probabilidad pequeña de mutar.

4. Reemplazamiento: Se repiten los 3 pasos anteriores hasta formar n hijos y poder reemplazar la población.

Con este proceso se obtiene una generación. El número de generaciones así como el tamaño de la población se fijan antes de iniciar con el algoritmo. En la Figura 5.1 podemos ver el diagrama de los pasos mencionados.

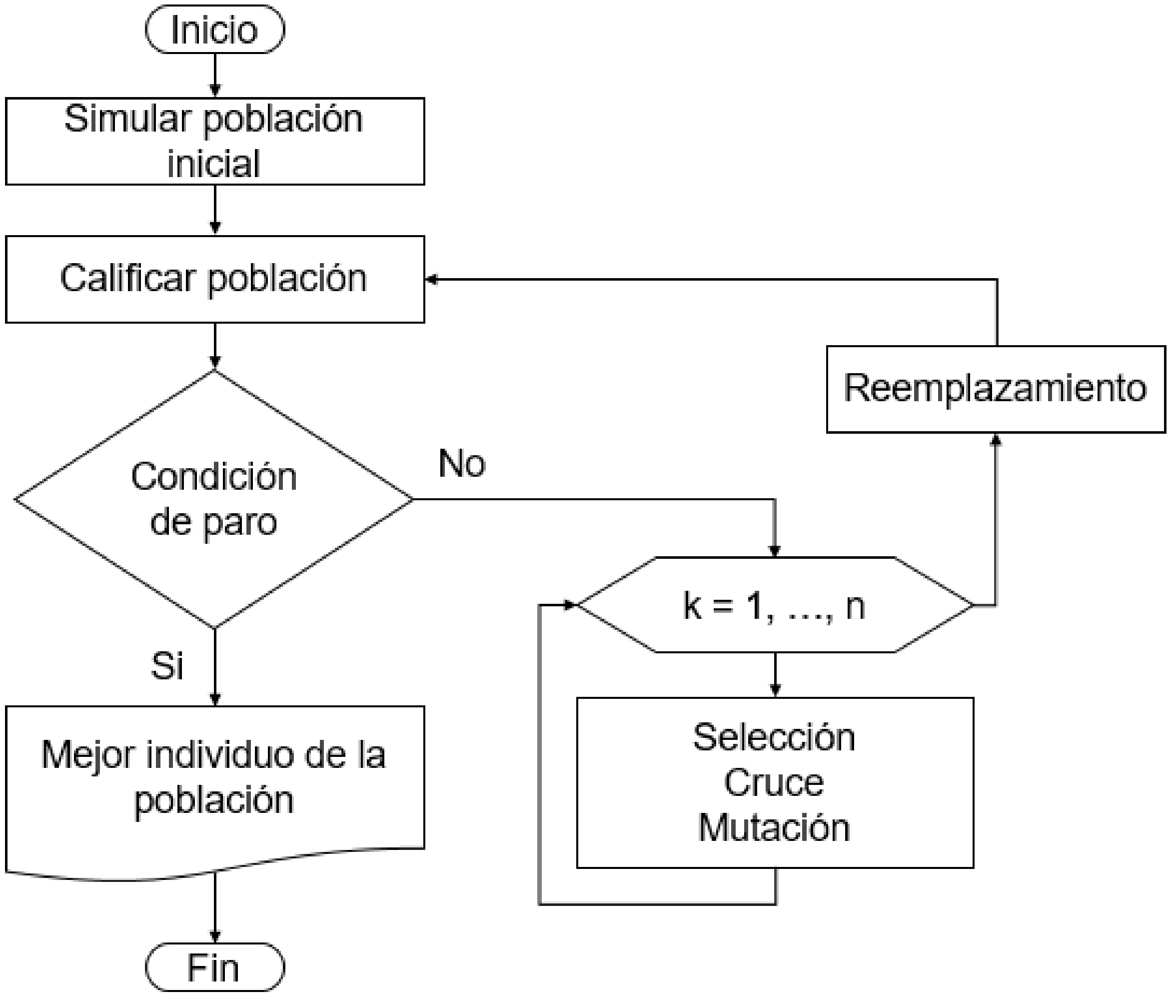


Figura 5.1: *Se muestra el diagrama de flujo que se sigue en el Algoritmo Genético.*

En la siguiente sección explicaremos cómo encontramos una buena asignación utilizando el AG. Cabe mencionar que Reeves y Rowe en su libro *Genetic Algorithms: Principles and Perspectives* [12], nos indican que se puede generar una nueva población haciendo el cruce y la mutación o utilizando sólo una de ellas. En nuestro caso, la estrategia que seguimos fue utilizar ambas.

5.1. Algoritmo Genético aplicado a los horarios

El objetivo principal de este proyecto es obtener una matriz con la asignación final de materias, profesores y horas. Para ello haremos uso del AG. A continuación vamos a definir los términos que utilizaremos:

- Asignación: Matriz de 3 columnas con la información de materias, profesores y horarios.
- Población: Conjunto de *n* asignaciones.
- Padre: Asignación seleccionada, de la población, para formar un hijo.

- Gen: Vector de 3 entradas (Materia, Profesor, Horario) con la información extraída de una asignación.
- Hijo: Asignación formada a partir de los genes de 2 padres.
- Probabilidad de mutación: Debe de ser un valor pequeño.
- Generación: Se dice que se tiene una generación cuando se ha repetido n veces el proceso para crear un hijo y se puede reemplazar la población.

Los pasos que seguimos para obtener la asignación son:

1. **Generar una población inicial**

   Se generan n asignaciones a partir del esqueleto simulado (ver Sección 4.10) y de las solicitudes pseudo-reales simuladas (ver Sección 4.5).

2. **Calificar cada asignación de la población**

   Cada asignación tiene 2 tipos de calificaciones:

   a) Por gen: Se califica cada gen de la asignación. Se premia con +5 si el profesor asignado es de tiempo completo. Se penaliza con -1 por cada asignación que pudo haber tenido un profesor de tiempo completo y tiene un profesor de asignatura. Para tener una calificación diferente para cada grupo, sumamos a cada gen una $\varepsilon \in [0, 0.1]$.

   b) Global: Se califica la asignación completa. Se penaliza con -1 por cada grupo en el esqueleto sin profesor. Se penaliza con -10 por cada materia pedida por algún profesor de tiempo completo y no se le asignó. Se suma el promedio de las calificaciones por gen.

   Nota: Si el número máximo de asignaciones es 2 y un profesor pidió 3 o más materias pero sólo se le asignó 1, entonces se penaliza una materia. Si se le asignaron, 2 no hay penalización.

3. **Ordenar de acuerdo a la calificación**

   Los genes de cada asignación se ordenan de menor a mayor calificación. Porque se quiere elegir con mayor probabilidad los genes con mejor calificación.

   Las asignaciones se ordenan de menor a mayor calificación. Porque se quiere elegir con mayor probabilidad las asignaciones con mejor calificación.

4. **Elegir 2 padres**

   Los padres se eligen con probabilidad: $\mathbb{P}(\text{elegir la asignación } i \text{ ya ordenada}) = \dfrac{2i}{n(n+1)}$, donde i es la posición de la asignación con respecto a su calificación.

5. **Elegir un gen**

   Primero de elige al azar un padre (ambos tienen probabilidad 0.5). Una vez que se eligió un padre, seleccionar un gen: $\mathbb{P}(\text{elegir el gen } i \text{ ya ordenado}) = \dfrac{2i}{g(g+1)}$, donde *i* es la posición del gen en la asignación con respecto a su calificación y *g* es el número de genes que tiene la asignación.

6. **Mutación**

   Se simula un número aleatorio, si ese número es menor a *prob_mutacion*, entonces el gen tiene una mutación. Si un gen muta, entonces se elige un gen de las solicitudes pseudo-reales y se intercambia por el gen previamente seleccionado.

7. **Agregar gen**

   Una vez definido el gen, éste se agrega al hijo.

8. **Ajustar información**

   Se quita la información a los padres, del profesor en el gen elegido:

   - A esa hora y con esa materia para evitar que se elijan genes repetidos para el hijo. Esto puede ocurrir por ejemplo cuando ambos padres tienen el mismo gen.
   - A esa hora porque los profesores no pueden impartir más de una clase a la misma hora.
   - Con esa materia porque no se les puede asignar la misma materia en diferentes horarios.
   - Cualquier materia a cualquier hora cuando el profesor ya tiene el número máximo de materias asignadas.

9. **Repetir 5 - 8**

   Repetir los pasos 5 al 8 hasta que uno de los padres se quede sin genes.

10. **Añadir genes**

    Agregar al hijo los genes del padre que aún tiene información.

11. **Repetir 4 - 10**

    Repetir los pasos 4 al 10 *n* veces para poder formar una generación.

12. **Reemplazar población**

    Reemplazar a la población con la que se formó la generación.

13. **Repetir 2 - 12**

    Repetir los pasos 2 al 12 hasta completar el número de generaciones deseadas.

14. **Asignación final**

    Definir la asignación final como el hijo mejor calificado de la última generación.

Algunas notas que se deben de considerar en la asignación final:

- Algunos profesores se les asignaron 2 cálculos
- Hay profesores que ya no imparten clases en la Facultad (nota en la sección de los profesores)

- No se asignaron todos los grupos simulados en el esqueleto.

5.2. Resultados del Algoritmo Genético

En esta sección presentamos los resultados obtenidos con el AG. Para obtener la matriz con la asignación final se simularon 6 generaciones (población inicial más 5 reemplazamientos). El tamaño de la población para cada generación es 10. El número de genes varía dependiendo de cada asignación.

En la Figura 5.2 vemos las calificaciones de la mejor asignación por generación. Observamos que la mejora en la calificación es considerable de la población inicial a la segunda generación. Notamos que la calificación del mejor elemento de la generación 5 fue menor al de la generación 4.

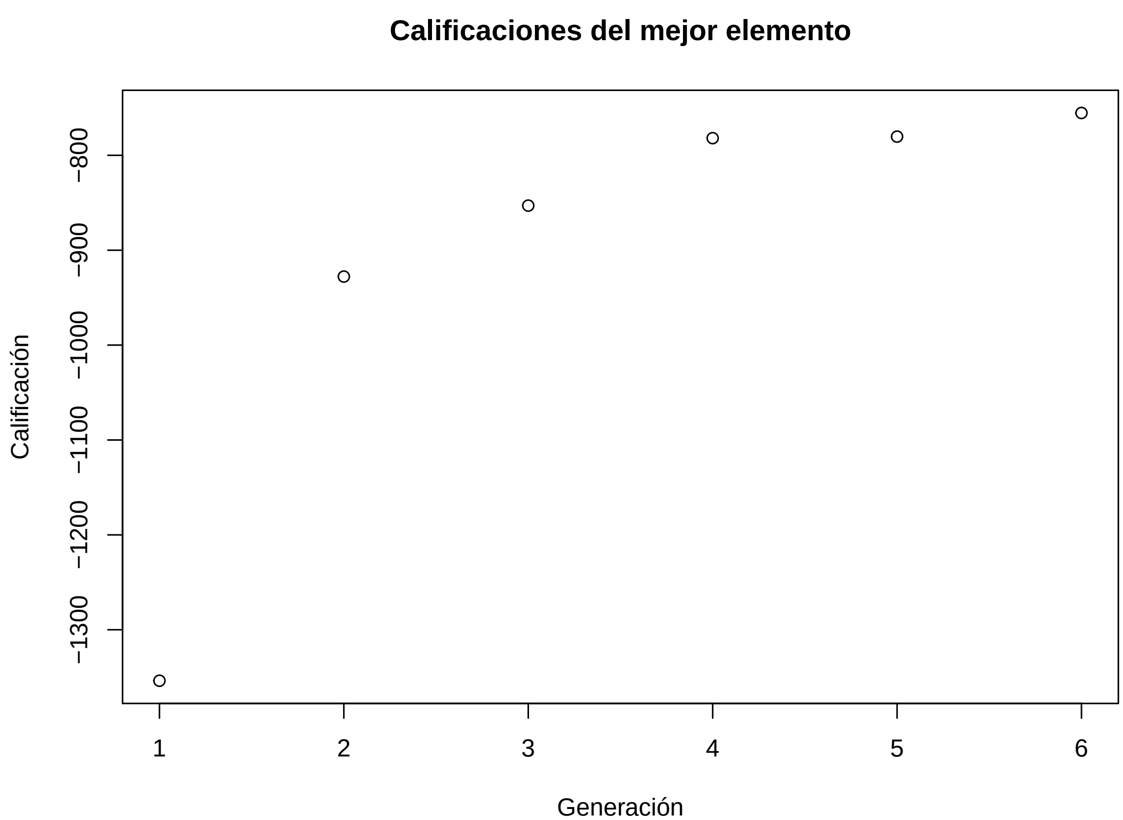


Figura 5.2: *Se muestran las calificaciones de la mejor asignación por generación. Se observa una mejora considerable en la calificación de la generación 1 a la 2.*

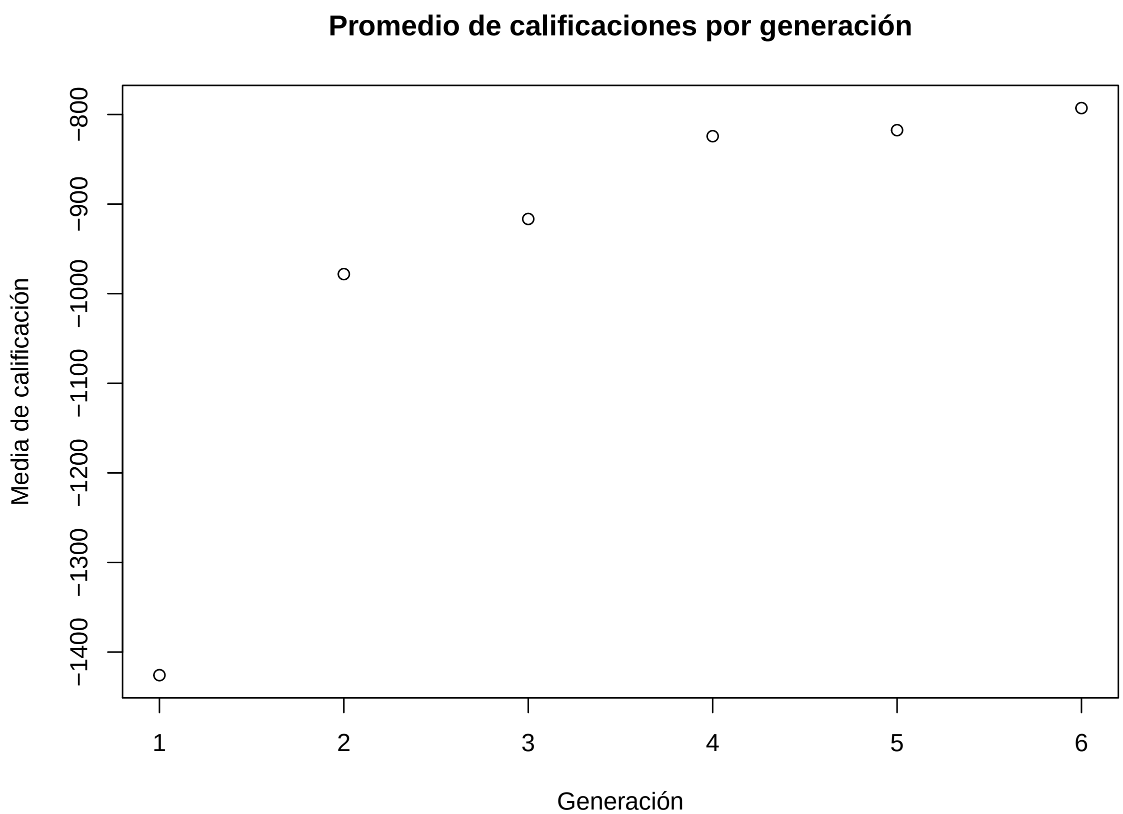


Figura 5.3: *Se muestra el promedio de las calificaciones de las asignaciones por generación. Se observa una mejora considerable en la calificación de la generación 1 a la 2.*

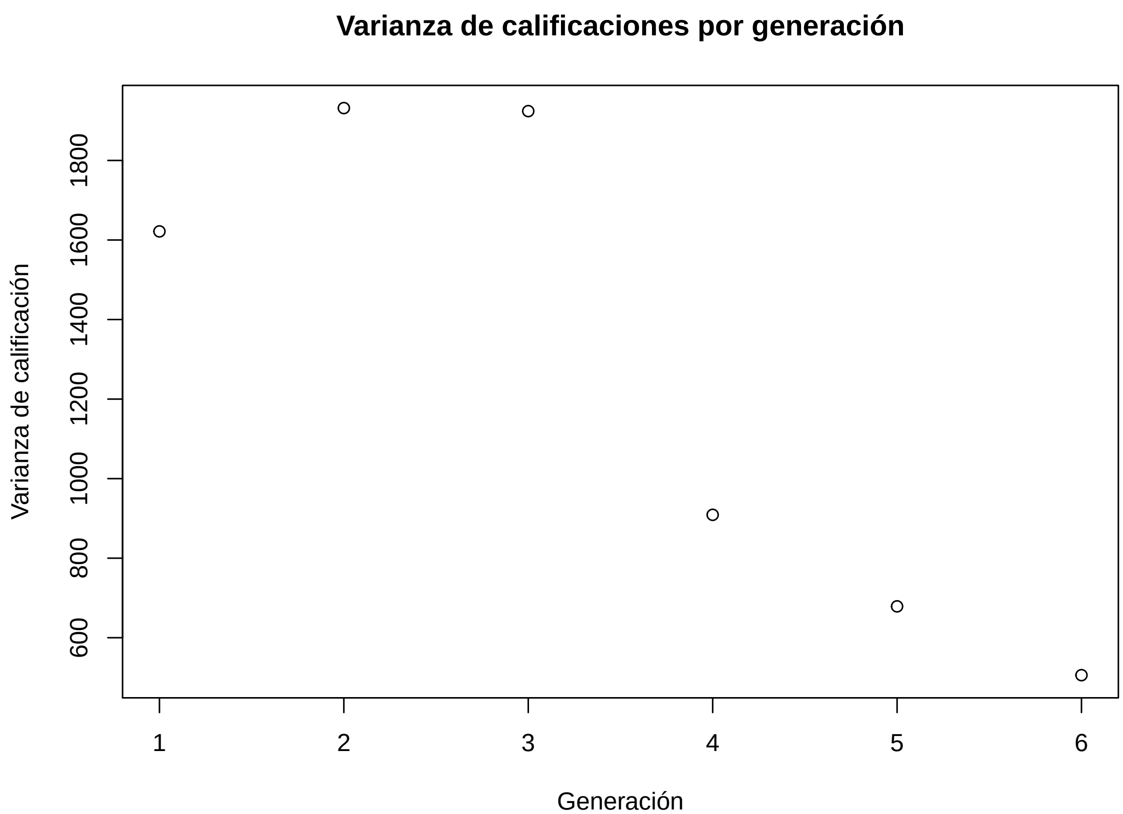


Figura 5.4: *Se muestra la varianza de las calificaciones de las asignaciones por generación. Se observa una disminución considerable de la generación 3 a la 4.*

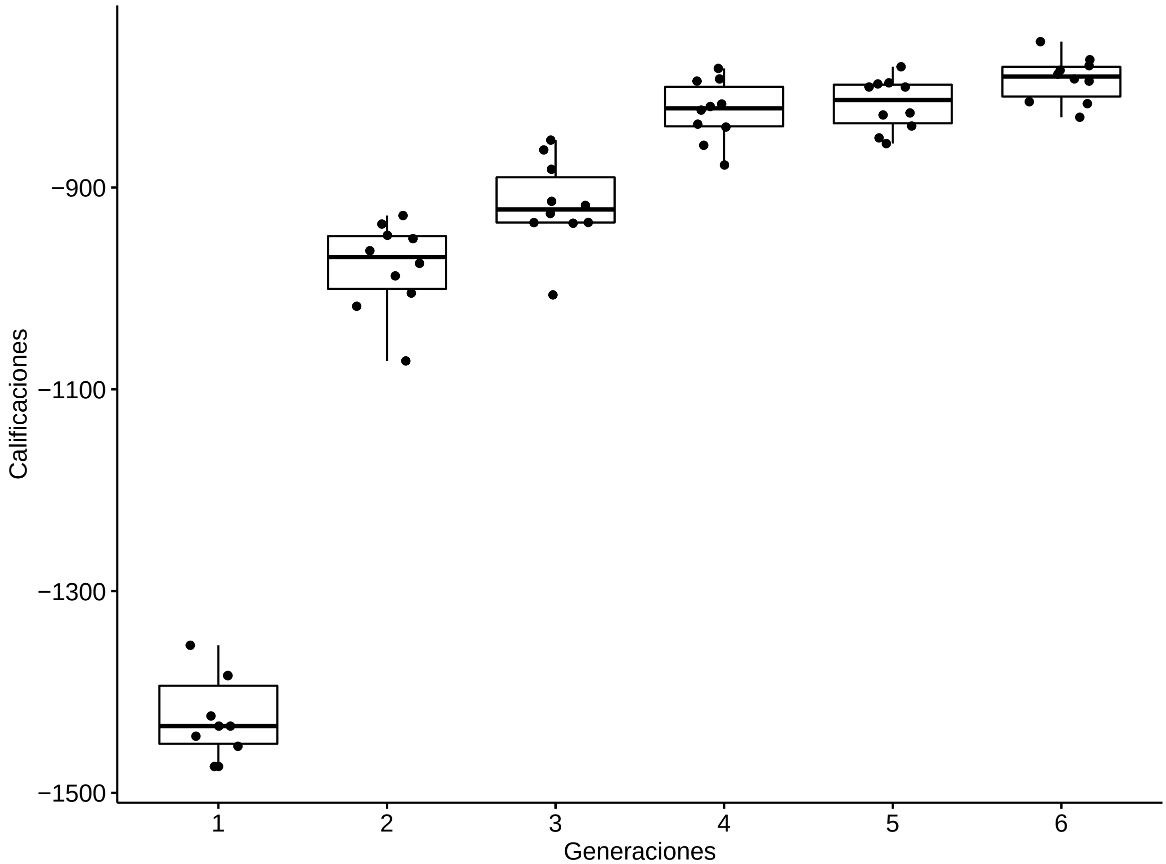


Figura 5.5: *Se muestra la gráfica de caja de las calificaciones de las asignaciones por generación. Se observa una mejora considerable en la calificación de la generación 1 a la 2. Los puntos respresentan las calificaciones de cada asignación por generación.*

En un esqueleto simulado para el 2020-2 se tuvieron 1018 grupos. De éstos se asignaron 612. Encontramos que 262 grupos sin asignación correspondían a materias optativas. Prácticamente $\frac{1}{4}$ de los grupos no asignados son de materias optativas. Ésto se debe a que se pueden asignar *num_max_asig* materias obligatorias a los profesores por lo que ya no se les asignan las optativas. En la matriz de solicitudes puede haber grupos que no están en la asignación final. Ésto por el cruce de los padres.

$$\begin{aligned}
\frac{612}{1018} &= 60.11\,\%\text{grupos asignados}\\
\frac{262}{1018} &= 25.73\,\%\text{grupos de optativas sin asignación}\\
\frac{1018-(612+262)}{1018} &= \frac{144}{1018} = 14.14\,\%\text{grupos de obligatorias sin asignación}
\end{aligned}$$

	mat_esq	esq_asig_fin	gpos_sin_asig	dif_rel
Series de Tiempo	2	0	2	1.00000000
Teoría de Redes	5	0	5	1.00000000
Cálculo de las Variaciones	1	0	1	1.00000000
Historia de las Matemáticas I	1	0	1	1.00000000
Teoría de los Conjuntos II	2	0	2	1.00000000
Econometría I	1	0	1	1.00000000
Análisis de Algoritmos	1	0	1	1.00000000
Lenguajes de Programación y sus Paradigmas	1	0	1	1.00000000
Programación Lineal	3	0	3	1.00000000
Teoría de las Decisiones	2	0	2	1.00000000
Seminario Filosofía de las Matemáticas	1	0	1	1.00000000
Geometría Diferencial II	1	0	1	1.00000000
Procesos Estocásticos II	1	0	1	1.00000000
Teoría de los Números II	1	0	1	1.00000000
Teoría de la Medida I	3	0	3	1.00000000
Temas Selectos de Investigación de Operaciones	1	0	1	1.00000000
Geometría Proyectiva	2	0	2	1.00000000
Auditoría Actuarial	1	0	1	1.00000000
Control Estadístico de la Calidad	1	0	1	1.00000000
Seminario de Ciencia y Sociedad I	2	0	2	1.00000000
Seminario de Ciencia y Sociedad II	1	0	1	1.00000000

Figura 5.6: *Optativas sin grupos asignados*

```
dim poblacion[[ 1 ]] =  654 6
dim poblacion[[ 2 ]] =  661 6
dim poblacion[[ 3 ]] =  659 6
dim poblacion[[ 4 ]] =  650 6
dim poblacion[[ 5 ]] =  654 6
dim poblacion[[ 6 ]] =  657 6
dim poblacion[[ 7 ]] =  648 6
dim poblacion[[ 8 ]] =  656 6
dim poblacion[[ 9 ]] =  659 6
dim poblacion[[ 10 ]] =  659 6
```

```
dim poblacion[[ 1 ]] =  563 6
dim poblacion[[ 2 ]] =  530 6
dim poblacion[[ 3 ]] =  569 6
dim poblacion[[ 4 ]] =  559 6
dim poblacion[[ 5 ]] =  531 6
dim poblacion[[ 6 ]] =  567 6
dim poblacion[[ 7 ]] =  532 6
dim poblacion[[ 8 ]] =  563 6
dim poblacion[[ 9 ]] =  522 6
dim poblacion[[ 10 ]] =  555 6
```

Figura 5.7: *Número de genes en generaciones 1 y 2*

Sabemos que la asignación final es el mejor elemento de la última generación. En la Tabla 5.1 presentamos una submatriz de la asignación final. Cabe aclarar que los datos se ordenaron con respecto a la materia (en orden alfabético) y por hora (de menor a mayor). La matriz completa se puede ver el el Apéndice D. Dicha matriz tiene 606 grupos asignados, cada uno con materia, profesor y horario correspondiente. En el semestre 2020-1 se tuvieron 747 grupos reales.

	Materia	Profesor	Horario
1	Administración Actuarial del Riesgo	Ricardo Villegas Azcorra	7
109	Análisis Numérico	Úrsula Xiomara Iturrarán Viveros	10
118	Cálculo Diferencial e Integral I	Javier Fernández García	7
153	Cálculo Diferencial e Integral III	Javier Fernández García	11
161	Cálculo Diferencial e Integral IV	Héctor Méndez Lango	10
442	Modelos de Supervivencia y de Series de Tiempo	Margarita Elvira Chávez Cano	10
444	Modelos de Supervivencia y de Series de Tiempo	Rubén Ugalde Franco	17
445	Modelos no Paramétricos y de Regresión	Margarita Elvira Chávez Cano	9
448	Modelos no Paramétricos y de Regresión	Lizbeth Naranjo Albarrán	11
449	Modelos no Paramétricos y de Regresión	Jaime Vázquez Alamilla	12
470	Probabilidad I	Jaime Vázquez Alamilla	8
476	Probabilidad I	Bibiana Obregón Quintana	14
496	Procesos Estocásticos I	Sergio Iván López Ortega	15
603	Variable Compleja I	Carisa Cano Figueroa	13

Tabla 5.1: *Se muestra una submatriz de la asignación final. Cada renglón tiene la información de un grupo con una materia, profesor y horario asignado.*

Capítulo 6

Conclusiones

La división que se hizo de los datos es estadísticamente adecuada.

Se encontró que el AG es una buena opción para solucionar este problema de maximización.

Este trabajo apoya las necesidades de los alumnos de la Facultad.

Con el fin de encontrar más posibles aplicaciones del procedimiento realizado en este trabajo, se buscaron diferentes páginas de horarios en distintas facultades de la UNAM y de otras universidades. No se pudieron encontrar páginas con todas las características que tienen las páginas de la Facultad. Debido a ésto, no se podría aplicar el procedimiento completo descrito en este trabajo.

Si se tiene información del número de alumnos por materia de varios semestres es posible implementar los algoritmos hechos para la creación de esqueletos (con el modelo de mezcla de normales) y la asignación final (con el algoritmo genético). Es decir lo único que no se podría utilizar es la aplicación *SelectorGadget*.

Algunas de las páginas que se encontraron son las siguientes:

- *Facultad de Ingeniería UNAM:* En la siguiente página web se puede seleccionar una materia y buscar la información de ella del semestre en curso, no se puede acceder a información de semestres anteriores y no se tiene alguna estructura para buscar de manera automática los datos.

  `https://www.ssa.ingenieria.unam.mx/horarios.html`

  Una vez que se ingresa a la materia, se puede encontrar información del salón, horario, cupo y vacantes, se podría obtener el número de alumnos inscritos al restar el cupo del número de vacantes, pero al no tener información de semestres anteriores en todo momento, la recopilación de información tardaría años.

- *Universidad La Salle:* Se encontró que las páginas tienen una cierta estructura y también se tiene la información del número de alumnos inscritos por materia pero los archivos son pdf por lo que no se puede utilizar la aplicación *SelectorGadget* para obtener la información.

  `https://cienciasquimicas.lasalle.mx/wp-content/uploads/2020/08/QFB-291.`

pdf

https://cienciasquimicas.lasalle.mx/wp-content/uploads/2020/08/QFB-391.pdf

https://cienciasquimicas.lasalle.mx/wp-content/uploads/2020/08/QFB-991.pdf

Sólo se encontró la misma estructura en las otras carreras de la Facultad, por lo que se puede ajustar el programa realizado en este trabajo para ellas. Algunas consideraciones que se deberían de tomar en cuenta son por ejemplo que las materias impartidas en los laboratorios duran más de una hora, no todas las materias se imparten todos los días, existen varias materias que no duran horas enteras. A continuación se presentan algunos ejemplos:

- *Biología:*

  http://www.fciencias.unam.mx/docencia/horarios/20172/181/1601

- *Ciencias de la Tierra:*

  http://www.fciencias.unam.mx/docencia/horarios/20182/1439/1318

- *Física:*

  http://www.fciencias.unam.mx/docencia/horarios/20191/1081/830

- *Física Biomédica:*

  http://www.fciencias.unam.mx/docencia/horarios/20192/2016/1735

- *Manejo Sustentable de Zonas Costeras:*

  http://www.fciencias.unam.mx/docencia/horarios/20181/1262/386

Apéndice A

Materias agrupadas

Vemos las materias que se actualizaron o cambiaron de nombre. Las negritas son los nombres y número de materia que tenían cuando el vector tenía 335 materias.

- Administración **(1)** -> Administración Actuarial (148) -> **Administración Actuarial del Riesgo** (288)
- Seminario de Inteligencia Artificial **(3)** -> **Recuperación y Búsqueda de Información en Textos** (257)
- **Seminario de Aplicaciones a las Ciencias Sociales y Administrativas (4)** -> Administración de Empresas de Software (258) -> Riesgo Tecnológico (278) -> Temas Selectos de Ingeniería de Software A (192)
- Probabilidad y Estadística **(5)** -> **Probabilidad I** (60)
- **Mecánica Vectorial (6)** -> Cálculo Tensorial (248)
- Matemáticas Avanzadas de la Física **(12)** -> **Funciones Especiales y Transformadas Integrales** (53) -> Análisis de Fourier I (208) -> Análisis de Fourier II (231) -> Introducción a las Funciones Recursivas y Computabilidad (224)
- **Mecánica Analítica (13)** -> Introducción Matemática a la Mecánica Celeste (119)
- **Física Computacional (15)** -> Supercómputo (195)
- **Teoría de Gráficas (33)** -> Teoría de las Gráficas II (147)
- Graficas y Juegos **(36)** -> **Introducción a las Matemáticas Discretas** (311)
- Estadística I **(41)** -> **Inferencia Estadística** (300)
- Análisis de Redes **(44)** -> **Teoría de Redes** (152)
- Bases de Datos **(50)** -> Formación Científica I (330) -> Sistemas Manejadores de Bases de Datos (106) -> Sistemas de Bases de Datos (123) -> Grandes Bases de Datos (169) -> Fundamentos de Bases de Datos (241) -> Almacenes y Minería de Datos (269) -> **Manejo de Datos** (301) -> Programación II (51)

- **Análisis Numérico (54)** -> Análisis Numérico II (161) -> Temas Selectos de Análisis Numérico (321)
- **Seminario sobre Enseñanza de las Matemáticas I (56)** -> Seminario de Filosofía de la Ciencia I (118) -> Didáctica de las Matemáticas (319)
- Estadística II **(59)** -> **Modelos no Paramétricos y de Regresión** (284) -> Análisis de Regresión (113)
- Teoría de la Computación **(67)** -> **Autómatas y Lenguajes Formales** (240)
- Matemáticas Discretas **(68)** -> **Estructuras Discretas** (220)
- Programación I **(69)** -> **Programación** (287)
- **Procesos Estocásticos I (70)** -> Procesos Estocásticos (159)
- **Seminario de Geometría A (73)** -> Álgebra Geométrica (207) -> Geometría Algebraica II (209)
- Fianzas **(78)** -> Matemáticas Actuariales del Seguro de Daños (79) -> **Matemáticas Actuariales para Seguro de Daños, Fianzas y Reaseguro** (SN) -> Matemáticas Actuariales para Seguro de Daños (297) -> Reaseguro (98) -> Reaseguro Financiero (127)
- Teoría de Juegos I (143) -> **Teoría de Juegos en Economía (80)**
- Finanzas II **(82)** -> **Métodos Cuantitativos en Finanzas** (298)
- Seminario de Aplicaciones Actuariales I (303) -> Seminario de Matemáticas Actuariales Aplicadas **(85)** -> Seminario de Aplicaciones Actuariales II (323) -> **Seminario de Aplicaciones Actuariales** (142) -> /Seminario de Aplicaciones Actuariales I/Seminario de Estadística I (333) -> Seminario de Probabilidad A (211) -> Teoría de la Medida II (313)
- Finanzas I **(86)** -> **Mercados Financieros y Valuación de Instrumentos** (306) -> Valuación de Opciones (128)
- Problemas Socio-Económicos de México **(87)** -> **Análisis del México Contemporáneo** (275) -> México: Nación Multicultural (176)
- Formación Científica II **(88)** -> **Economía** (304) -> Economía I (88)
- Productos Financieros Derivados I **(91)** -> Productos Financieros Derivados II (184) -> **Productos Financieros Derivados** (326)
- Economía II **(93)** -> **Temas Selectos de Economía** (307) -> Econometría II (233)
- Demografía I **(94)** -> Demografía II (97) -> **Demografía** (289) -> Demografía Avanzada (308)
- Introducción a Ciencias de la Computación I **(95)** -> Introducción a Ciencias de la Computación II (101) -> **Introducción a Ciencias de la Computación** (222) -> Estructuras de Datos (234) -> Robótica (268)
- Arquitectura de Computadoras **(102)** -> **Organización y Arquitectura de Computadoras** (245)

- Análisis de Algoritmos I **(103)** -> Análisis de Algoritmos II (205) -> **Análisis de Algoritmos** (243)
- **Lenguajes de Programación (104)** -> Lenguajes de Programación y sus Paradigmas (247) -> Semántica y Verificación (214)
- Seminario de Ciencias de la Computación A (254) -> **Seminario de Ciencias de la Computación** (SN) -> Seminario de Ciencias de la Computación B (263) -> Seminario de Temas Selectos de Computación **(105)** -> Seminario de Aplicaciones de Cómputo (133) -> Seminario de Computación Teórica (162) -> Seminario de Aplicaciones de Cómputo II (191) -> Seminario de Sistemas para Cómputo B (217) -> Seminario de Computación Teórica II (228) -> Seminario de Sistemas para Cómputo A (164) -> Administración de Sistemas Unix/Linux (282) -> Sistemas de Información Geográfica (274) -> Métodos Formales (291)
- Principios de Computación Distribuida (190) -> Computación Concurrente (259) -> **Computación Distribuida** (252) -> **(202)**
- **Animación por Computadora** (255) -> **(203)**
- Seminario de Programación **(107)** -> **Modelado y Programación** (246) -> Diseño y Programación Orientada a Objetos (168) -> Programación Funcional y Lógica (196) -> Programación de Dispositivos Móviles (277) -> Programación Declarativa (296)
- Análisis Lógico **(108)** -> **Lógica Computacional** (244) -> Lógica Computacional II (251) -> Lógicas no Clásicas (166)
- Diseño de Sistemas Digitales **(130)** -> **Diseño de Interfaces de Usuario** (272) -> Diseño de interfaces (167)
- Seminario de Inteligencia Artificial II (163) -> Reconocimiento de Patrones (264) -> **Reconocimiento de Patrones y Aprendizaje Automatizado** (281) -> Seminario de Temas Selectos de Computación II **(132)** -> Computación Cuántica I (267) -> Computación Cuántica II (279) -> Sistemas Expertos (198) -> Razonamiento Automatizado (292)
- **Seminario Filosofía de las Matemáticas (135)** -> Seminario de Filosofía de la Ciencia II (138) -> Seminario de Filosofía de la Ciencia III (155) -> Seminario de Filosofía de la Ciencia IV (146)
- Estadística III (139) -> **Modelos de Supervivencia y de Series de Tiempo** (285) -> Series de Tiempo **(41)**
- **Seminario Matemáticas Aplicadas I (144)** -> Seminario de Cálculo de Formas Diferenciales (273)
- Seminario de Investigación de Operaciones **(160)** -> **Temas Selectos de Investigación de Operaciones** (305)
- Temas Selectos de Ingeniería de Software B **(165)** -> Temas Selectos de Ingeniería de Software A (192) -> Tecnologías para Desarrollos en Internet (265) -> **Ingeniería de Software II** (283) -> Patrones de Diseño de Software (294)
- Diseño de Experimentos **(177)** -> **Seminario de Estadística I** (324)

- **Seminario de Topología B (179)** -> Topología Diferencial II (232)
- **Mercadotecnia de Seguros (183)** -> Contabilidad de Seguros (290)
- **Graficación por Computadoras (186)** -> Visualización (188) -> Geometría Computacional (213) -> Visión Por Computadora (293)
- Seminario de Ciencias Computacionales **(189)** -> **Taller de Herramientas Computacionales** (312) -> Sistemas Dinámicos Computacionales I (276) -> Lingüística Computacional (227) -> Herramientas de Computación para las Ciencias (229) -> Algoritmos de Apareamiento de Cadenas (286)
- Redes Neuronales y Autómatas Celulares **(193)** -> **Redes Neuronales** (302)
- Procesos Paralelos y Distribuidos **(194)** -> **Algoritmos Paralelos** (270)
- Algoritmos Genéticos **(197)** -> **Cómputo Evolutivo** (280)
- Simulación y Control **(201)** -> **Control Estadístico de la Calidad** (210)
- Introducción a la Criptografía **(262)** -> **Criptografía y Seguridad** (271)
- **Seminario de Apoyo a la Titulación en Ciencias de la Computación** (SN) -> Seminario de Apoyo a la Titulación en Ciencias de la Computación A **(315)** -> Seminario de Apoyo a la Titulación en Ciencias de la Computación B (316)
- **Seminario de Apoyo a la Titulación en Matemáticas** (SN) -> Seminario de Apoyo a la Titulación en Matemáticas A **(310)** -> Seminario de Apoyo a la Titulación en Matemáticas B (317)

Apéndice B

Resultados útiles

Definición B.1. ***Estimador máximo verosímil de*** λ

Sean $X_1, X_2, \ldots, X_n$ *una muestra aleatoria de una población con función de densidad de probabilidad Poisson*(λ)*. Su función de densidad es:*

$$f(x) = \mathrm{e}^{-\lambda} \frac{\lambda^x}{x!} \tag{B.1}$$

$$
\begin{aligned}
\mathscr{L}(X_1, X_2, \ldots, X_n; \lambda) &= \prod_{i=1}^{n} \left(\mathrm{e}^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \right) \\
&= \mathrm{e}^{-n\lambda} \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \\
\textit{Sacamos ln} & \\
ln\mathscr{L}(X_1, X_2, \ldots, X_n; \lambda) &= -n\lambda + \sum_{i=1}^{n} x_i ln\lambda - ln \prod_{i=1}^{n} x_i! \\
\textit{Derivamos con respecto a } \lambda & \\
\frac{\partial}{\partial \lambda} ln\mathscr{L}(\underline{X}; \lambda) &= -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} \\
\textit{Igualamos a cero} & \\
-n + \frac{\sum_{i=1}^{n} x_i}{\lambda} &= 0 \\
\Rightarrow & \\
\textit{Despejamos } \lambda & \\
\hat{\lambda} &= \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \\
\textit{Derivamos otra vez} & \\
\frac{\partial^2}{\partial \lambda} ln\mathscr{L}(\underline{X}; \lambda) &= -\frac{\sum_{i=1}^{n} x_i}{\lambda^2} < 0 \\
\therefore \hat{\lambda} &= \bar{x} \quad \textit{es el estimador máximo verosímil}
\end{aligned}
$$

Apéndice C

Abreviaturas

ABREVIATURA	SIGNIFICADO
AG	Algoritmo Genético
CdC	Ciencias de la Computación
Facultad	Facultad de Ciencias de la UNAM
FES	Facultad de Estudios Superiores
ITAM	Instituto Tecnológico Autónomo de México
MatAp	Matemáticas Aplicadas
TC	Tiempo Completo
UNAM	Universidad Nacional Autónoma de México
URL	Uniform Resource Locator
a	b

Tabla C.1: *Abreviaturas*

Apéndice D

Ejemplo de asignación final

Sabemos que la asignación final es el mejor elemento de la última generación. Éste elemento es la matriz que presentamos a continuación. Cabe aclarar que los datos se ordenaron con respecto a la materia (en orden alfabético) y por hora (de menor a mayor). La matriz tiene 606 grupos asignados, cada uno con materia, profesor y horario correspondiente.

	Materia	**Profesor**	**Horario**
1	Administración Actuarial del Riesgo	Ricardo Villegas Azcorra	7
2	Administración Actuarial del Riesgo	María Patricia Luna Díaz	9
3	Administración Actuarial del Riesgo	Oscar Lucio Cano Vaca	20
4	Administración de Riesgos	José Antonio Reyes León	8
5	Administración de Riesgos Financieros	Jesús David Gómez Téllez	8
6	Álgebra Lineal I	Karina García Buendía	7
7	Álgebra Lineal I	Angel Vázquez Badillo	7
8	Álgebra Lineal I	Irvin Arellano Rosas	7
9	Álgebra Lineal I	Alma Violeta García López	8
10	Álgebra Lineal I	César Alejandro Rincón Orta	8
11	Álgebra Lineal I	Gabriela Campero Arena	9
12	Álgebra Lineal I	Francisco Marmolejo Rivas	10
13	Álgebra Lineal I	Adolfo Guillot Santiago	10
14	Álgebra Lineal I	Jorge Luis Arocha Pérez	10
15	Álgebra Lineal I	Hugo Arizmendi Peimbert	10
16	Álgebra Lineal I	Gustavo Amilcar Saldaña Moncada	10
17	Álgebra Lineal I	Clotilde García Villa	11
18	Álgebra Lineal I	María del Carmen Heréndira Gómez Laveaga	12
19	Álgebra Lineal I	William José Gallardo	14
20	Álgebra Lineal I	José Patricio Sánchez Hernández	16
21	Álgebra Lineal I	Víctor Manuel Ávila Baez	20

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
22	Álgebra Lineal I	Raúl Bartolo Martínez	20
23	Álgebra Lineal I	Francisco de Jesús Rivera Torres	21
24	Álgebra Lineal II	Carlos Alberto Serrato Hernández	7
25	Álgebra Lineal II	Lorena Morales Callejas	8
26	Álgebra Lineal II	Andres Barei Bueno	8
27	Álgebra Lineal II	Gustavo Amilcar Saldaña Moncada	9
28	Álgebra Lineal II	Guillermo Javier Francisco Sienra Loera	11
29	Álgebra Lineal II	Juan Morales Rodríguez	13
30	Álgebra Lineal II	Alejandro Alvarado García	14
31	Álgebra Lineal II	Leobardo Fernández Román	17
32	Álgebra Moderna I	Mindy Yaneli Huerta Pérez	7
33	Álgebra Moderna I	Edith Corina Sáenz Valadez	9
34	Álgebra Moderna I	Hugo Alberto Rincón Mejía	11
35	Álgebra Moderna I	Sergio Hiroki Koike Quintanar	15
36	Álgebra Moderna II	José Gabriel Ocampo Márquez	7
37	Álgebra Moderna II	Bertha María Tomé Arreola	10
38	Álgebra Moderna II	Edith Corina Sáenz Valadez	11
39	Álgebra Moderna II	Alberto Alcalá Álvarez	17
40	Álgebra Moderna III	Valente Santiago Vargas	9
41	Álgebra Superior I	Alejandra Osiris Romero Juárez	7
42	Álgebra Superior I	Itzel Jeanne Riquelme Cherrier	7
43	Álgebra Superior I	Daniela Mariyet Terán Guerrero	7
44	Álgebra Superior I	Martha Takane Imay	10
45	Álgebra Superior I	Sergio Macías Álvarez	10
46	Álgebra Superior I	Francisco Larrión Riveroll	10
47	Álgebra Superior I	Elsa Puente Vázquez	10
48	Álgebra Superior I	Carmen Martínez-Adame Isaís	10
49	Álgebra Superior I	Alejandro Bravo Mojica	10
50	Álgebra Superior I	Rita Esther Zuazua Vega	10
51	Álgebra Superior I	Natalia Bárbara Mantilla Beniers	11
52	Álgebra Superior I	Juan Morales Jr Rodríguez	12

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
53	Álgebra Superior I	Gabriela Campero Arena	13
54	Álgebra Superior I	Ernesto Mayorga Saucedo	14
55	Álgebra Superior I	Natalia Bárbara Mantilla Beniers	15
56	Álgebra Superior I	Marcelino Perello Valls	16
57	Álgebra Superior I	Juan Manuel Martínez Nuño	19
58	Álgebra Superior I	Eugenia Marmolejo Rivas	20
59	Álgebra Superior II	Tania Eréndira Rivera Torres	7
60	Álgebra Superior II	Rolando Gómez Macedo	7
61	Álgebra Superior II	Luis Jesús Turcio Cuevas	8
62	Álgebra Superior II	Lilia Guadalupe Sánchez Terán	8
63	Álgebra Superior II	Carmen Martínez Adame Isais	9
64	Álgebra Superior II	Araceli Guzmán Tristán	10
65	Álgebra Superior II	Gerardo Camacho de la Rosa	10
66	Álgebra Superior II	Carmen Martínez Adame Isais	10
67	Álgebra Superior II	Juan Morales Rodríguez	10
68	Álgebra Superior II	Juan Morales Jr Rodríguez	11
69	Álgebra Superior II	Patricia Cortés Flores	11
70	Álgebra Superior II	Antonio Lascurain Orive	12
71	Álgebra Superior II	Mauricio Gabriel Medina Bárcenas	15
72	Álgebra Superior II	Ana Patricia Kuri González	17
73	Álgebra Superior II	José Pozo Martínez	19
74	Álgebra Superior II	Francisco Páez Pérez	19
75	Álgebra Superior II	Fernando García Rodríguez	19
76	Álgebra Superior II	Alejandro Dorantes Aldama	19
77	Álgebra Superior II	Israel Zamorano Romero	20
78	Análisis del México Contemporáneo	Reyna Pineda González	8
79	Análisis del México Contemporáneo	Silvia Alonso Reyes	9
80	Análisis del México Contemporáneo	Dora Evangelina Mendizabal García	12
81	Análisis Matemático I	Ricardo Alberto Weder Zaninovich	7
82	Análisis Matemático I	Rodrigo Jesús Hernández Gutiérrez	7
83	Análisis Matemático I	Roxana Wendoline Ruiz Aguilar	8
84	Análisis Matemático I	Alberto León Kushner Schnur	9
85	Análisis Matemático I	Roberto Méndez Rosas	10

La tabla continúa en la siguiente página

	Materia	**Profesor**	**Horario**
86	Análisis Matemático I	Sergey Antonyan	10
87	Análisis Matemático I	Vinicio Antonio Gómez Gutiérrez	10
88	Análisis Matemático I	Roberto Pichardo Mendoza	12
89	Análisis Matemático I	Julián Fernando Chagoya Saldaña	13
90	Análisis Matemático I	Grissel Santiago González	15
91	Análisis Matemático I	Manuel Eduardo Chacón Ochoa	16
92	Análisis Matemático I	Moises Soto Bajo	17
93	Análisis Matemático I	Alfredo Reyes Vázquez	18
94	Análisis Matemático I	Gerardo Gonzalez Robert	18
95	Análisis Matemático I	Juan Manuel de la Huerta Jiménez	20
96	Análisis Matemático II	Esteban Librado Hernández Escamilla	7
97	Análisis Matemático II	Manuel Jesús Falconi Magaña	9
98	Análisis Matemático II	Ana Meda Guardiola	10
99	Análisis Matemático II	Angel Manuel Carrillo Hoyo	11
100	Análisis Matemático II	Gerardo Sánchez Licea	13
101	Análisis Matemático II	Juan Rico Arvizu	16
102	Análisis Matemático II	Miguel Ángel Corona García	17
103	Análisis Matemático III	María de la Luz Jimena de Teresa de Oteyza	11
104	Análisis Matemático III	Magali Louise Marie Folch Gabayet	13
105	Análisis Multivariado	Sofía Villers Gómez	10
106	Análisis Numérico	Antonio Carrillo Ledesma	7
107	Análisis Numérico	Jaime Ayala Pérez	9
108	Análisis Numérico	Ursula Xiomara Iturrarán Viveros	10
109	Análisis Numérico	Úrsula Xiomara Iturrarán Viveros	10
110	Análisis Numérico	Pablo Barrera Sánchez	10
111	Análisis Numérico	Guilmer Ferdinand González Flores	11
112	Análisis Numérico	Mario Medina Torres	17
113	Análisis Numérico	Miguel Ángel Pérez León	18
114	Análisis Numérico	Miriam Sosa Díaz	19
115	Autómatas y Lenguajes Formales	Elisa Viso Gurovich	10
116	Autómatas y Lenguajes Formales	Lourdes del Carmen González Huesca	11

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
117	Cálculo Diferencial e Integral I	Eric Fabián Hernández Martínez	7
118	Cálculo Diferencial e Integral I	Javier Fernández García	7
119	Cálculo Diferencial e Integral I	Mónica De Nova Vázquez	9
120	Cálculo Diferencial e Integral I	Miguel Lara Aparicio	10
121	Cálculo Diferencial e Integral I	Joel García León	10
122	Cálculo Diferencial e Integral I	María Lourdes Velasco Arreguí	11
123	Cálculo Diferencial e Integral I	María de Lourdes Esteva Peralta	12
124	Cálculo Diferencial e Integral I	José Antonio Gómez Ortega	13
125	Cálculo Diferencial e Integral I	Edgar René Hernández Martínez	14
126	Cálculo Diferencial e Integral I	Luis Eduardo García Hernández	15
127	Cálculo Diferencial e Integral I	Roberto García Medina	16
128	Cálculo Diferencial e Integral I	Erick García Ramírez	16
129	Cálculo Diferencial e Integral I	Sergio César Alejandro Gutiérrez Guzmán	17
130	Cálculo Diferencial e Integral II	Felipe de Jesús Méndez Varela	7
131	Cálculo Diferencial e Integral II	Elena de Oteyza de Oteyza	7
132	Cálculo Diferencial e Integral II	David Meza Alcántara	8
133	Cálculo Diferencial e Integral II	María del Carmen Arrillaga Arjona	10
134	Cálculo Diferencial e Integral II	María Lourdes Velasco Arreguí	10
135	Cálculo Diferencial e Integral II	Sebastián Nájera Valencia	12
136	Cálculo Diferencial e Integral II	Fernando Brambila Paz	13
137	Cálculo Diferencial e Integral II	Emilio Cabrera Castro	14
138	Cálculo Diferencial e Integral II	Humberto Andrés Carrillo Calvet	14
139	Cálculo Diferencial e Integral II	Wilfrido Martínez Torres	15
140	Cálculo Diferencial e Integral II	José Santos	15
141	Cálculo Diferencial e Integral II	Roberto García Medina	15
142	Cálculo Diferencial e Integral II	Francisco Javier Torres Ayala	15
143	Cálculo Diferencial e Integral II	Sergio Iker Martínez Juárez	16
144	Cálculo Diferencial e Integral II	María Juana Linares Altamirano	17
145	Cálculo Diferencial e Integral II	Aarón Aparicio Hernández	17
146	Cálculo Diferencial e Integral II	Jesús López Estrada	18
147	Cálculo Diferencial e Integral II	Julio César Cedillo Sánchez	19
148	Cálculo Diferencial e Integral III	Claudio Francisco Nebbia Rubio	7

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
149	Cálculo Diferencial e Integral III	Alejandro Darío Rojas Sánchez	9
150	Cálculo Diferencial e Integral III	Pablo Alberto Lara Martínez	10
151	Cálculo Diferencial e Integral III	Adriana Vargas Quintero	10
152	Cálculo Diferencial e Integral III	Julio Martín Espinosa Casares	10
153	Cálculo Diferencial e Integral III	Javier Fernández García	11
154	Cálculo Diferencial e Integral III	Álvaro Reyes García	14
155	Cálculo Diferencial e Integral III	Javier Páez Cárdenas	14
156	Cálculo Diferencial e Integral III	Abelardo Vela Ponce de León	16
157	Cálculo Diferencial e Integral III	Adrián Zenteno Gutiérrez	20
158	Cálculo Diferencial e Integral IV	Emily Sánchez García	9
159	Cálculo Diferencial e Integral IV	Oscar Alberto Garrido Jiménez	10
160	Cálculo Diferencial e Integral IV	Luis Manuel Hernández Gallardo	10
161	Cálculo Diferencial e Integral IV	Héctor Méndez Lango	10
162	Cálculo Diferencial e Integral IV	Javier Páez Cárdenas	10
163	Cálculo Diferencial e Integral IV	Jefferson Edwin King Dávalos	11
164	Cálculo Diferencial e Integral IV	Carlos Arturo Vargas Guadarrama	12
165	Cálculo Diferencial e Integral IV	Carlos Gerardo Paniagua Ramírez	13
166	Cálculo Diferencial e Integral IV	Adriana Vargas Quintero	18
167	Cálculo Diferencial e Integral IV	Abelardo Vela Ponce de León	18
168	Cálculo Diferencial e Integral IV	Gabriel Gutiérrez García	19
169	Cálculo Diferencial e Integral IV	Emilio Cabrera Castro	20
170	Compiladores	Adrián Ulises Mercado Martínez	8
171	Complejidad Computacional	Francisco Hernández Quiroz	10
172	Conjuntos y Lógica	César Hernández Cruz	8
173	Conjuntos y Lógica	Mario Francisco Rosales González	11
174	Conjuntos y Lógica	Fernando Javier Nuñez Rosales	18
175	Contabilidad	Pedro Luis Soto Tejeda	7
176	Contabilidad	Guillermo Vega García	8
177	Contabilidad	María de los Ángeles Garduño Crespo	8
178	Contabilidad	Arely Palos Zepeda	17
179	Contabilidad	Carlos Orozco Rocha	19
180	Demografía	José Rubén Fernández Román	7

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
181	Demografía	María Teresa Velázquez Uribe	8
182	Demografía	Nina Castro Méndez	9
183	Demografía	Rosalba Jasso Vargas	11
184	Demografía	Alejandro Mina Valdés	17
185	Demografía	Ángel Jair Morales Eslava	18
186	Dinámica de Medios Deformables	Guillermo Monsivais Galindo	8
187	Dinámica de Medios Deformables	Roberto Velasco Segura	16
188	Econometría I	Luis Alejandro Aguilar Luna	8
189	Economía	Marco Antonio García Fernández	8
190	Economía	Monserrat Esquivel López	10
191	Economía	Julio Bernabe Franco Moreno	17
192	Economía	Jessika Dilhery Lucas Flores	18
193	Ecuaciones Diferenciales I	Sara Jacqueline Herrera Domínguez	7
194	Ecuaciones Diferenciales I	Raziel Zavaleta Rodríguez	7
195	Ecuaciones Diferenciales I	Luisa Márquez Rentería	8
196	Ecuaciones Diferenciales I	Mirella Ramírez Ramírez	8
197	Ecuaciones Diferenciales I	Nicolás González Boileau	8
198	Ecuaciones Diferenciales I	Laura Ortiz Bobadilla	8
199	Ecuaciones Diferenciales I	Renato Carlos Calleja Castillo	9
200	Ecuaciones Diferenciales I	Carlos García Azpeitia	10
201	Ecuaciones Diferenciales I	Rocío del Pilar Aguilar Benítez	10
202	Ecuaciones Diferenciales I	Araceli León Estrada	10
203	Ecuaciones Diferenciales I	Manuel Jesús Falconi Magaña	10
204	Ecuaciones Diferenciales I	Gustavo Cruz Pacheco	11
205	Ecuaciones Diferenciales I	Arturo Olvera Chávez	12
206	Ecuaciones Diferenciales I	Roxana Wendoline Ruiz Aguilar	13
207	Ecuaciones Diferenciales I	Adrián Ulises Soto Bañuelos	15
208	Ecuaciones Diferenciales I	María de Lourdes Hernández Campos	17
209	Ecuaciones Diferenciales I	Fidencio Galicia Rodríguez	18
210	Ecuaciones Diferenciales I	Jorge Andrés Rosas Ávila	20
211	Ecuaciones Diferenciales I	Blanca Angélica Gónzalez Morales	21
212	Ecuaciones Diferenciales II	Ernesto Rosales González	8
213	Ecuaciones Diferenciales II	Jorge Chávez Carlos	9
214	Ecuaciones Diferenciales II	Humberto Andrés Carrillo Calvet	11

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
215	Ecuaciones Diferenciales II	José Luis Navarro Urrutia	16
216	Ecuaciones Diferenciales II	María de Lourdes Esteva Peralta	17
217	Ecuaciones Diferenciales Parciales I	María de los Ángeles Sandoval Romero	10
218	Electromagnetismo I	Ricardo Méndez Fragoso	7
219	Electromagnetismo I	Eugenio Ley Koo	7
220	Electromagnetismo I	Alicia María Oliver y Gutiérrez	8
221	Electromagnetismo I	Miguel Ángel Monroy Rodríguez	8
222	Electromagnetismo I	Julio Javier Martinell Benito	10
223	Electromagnetismo I	Alejandro Reyes Coronado	10
224	Electromagnetismo I	Andrea Luisa Aburto Espina	10
225	Electromagnetismo I	José Alberto Flandes Mendoza	16
226	Electromagnetismo II	Luis Fernando Urrutia Rios	10
227	Electromagnetismo II	Enrique Moreno Méndez	10
228	Electromagnetismo II	Alejandro Reyes Coronado	16
229	Estadística Bayesiana	Ruth Selene Fuentes García	9
230	Estructuras Discretas	Laura Freidberg Gojman	10
231	Estructuras Discretas	Alma Rosario Arévalo Loyola	14
232	Fenómenos Colectivos	Catalina Elizabeth Stern Forgach	8
233	Fenómenos Colectivos	Edgar Alvarez Zauco	9
234	Fenómenos Colectivos	Iván Santamaría Holek	10
235	Fenómenos Colectivos	Luis Felipe del Castillo Dávila	10
236	Fenómenos Colectivos	Alan Joel Miralrio Pineda	10
237	Fenómenos Colectivos	Marcos Ley Koo	12
238	Física Computacional	Edgar Vázquez Luis	7
239	Física Computacional	Ramón Gustavo Contreras Mayen	14
240	Física Computacional	Federico Jesús Cázarez Bush	18
241	Física Estadística	Gerardo Carmona Ruiz	8
242	Física Estadística	Guillermo Ramírez Santiago	10
243	Física Estadística	Víctor Manuel Romero Rochín	10
244	Física Estadística	Carlos Ramírez Ramos	18
245	Funciones Especiales y Transformadas Integrales	Luis Antonio Dávalos Orozco	10
246	Funciones Especiales y Transformadas Integrales	Guillermo Monsivais Galindo	10

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
247	Funciones Especiales y Transformadas Integrales	Lucía Medina Gómez	11
248	Funciones Especiales y Transformadas Integrales	Osvaldo Alfonso Téllez Nieto	13
249	Funciones Especiales y Transformadas Integrales	Ramón Gustavo Contreras Mayen	15
250	Funciones Especiales y Transformadas Integrales	Edward Daniel Reyes Ramírez	18
251	Geometría Analítica I	Juan Flores Torres	7
252	Geometría Analítica I	Max Neumann Coto	8
253	Geometría Analítica I	Jorge Chávez Carlos	8
254	Geometría Analítica I	Rodrigo Jesús Hernández Gutiérrez	8
255	Geometría Analítica I	Iván Axell Gómez Ramos	9
256	Geometría Analítica I	Edgar Alvarez Zauco	10
257	Geometría Analítica I	Guillermo Ruiz Galván	10
258	Geometría Analítica I	Jesús Ángel Núñez Zimbrón	10
259	Geometría Analítica I	Federico Sánchez Bringas	10
260	Geometría Analítica I	Guillermo Javier Francisco Sienra Loera	10
261	Geometría Analítica I	Oscar Alfredo Palmas Velasco	12
262	Geometría Analítica I	Noel Jaramillo Arce	14
263	Geometría Analítica I	Adolfo Guillot Santiago	17
264	Geometría Analítica I	Jonathan Giovanni Gil Juárez	18
265	Geometría Analítica II	Álvaro Reyes García	7
266	Geometría Analítica II	Pablo Rosell González	7
267	Geometría Analítica II	Rebeca Trejo Luna	8
268	Geometría Analítica II	Carlos Hernández Garciadiego	8
269	Geometría Analítica II	Gilberto Bruno Pérez	8
270	Geometría Analítica II	Alejandra García García	10
271	Geometría Analítica II	Francisco Manuel Barrios Paniagua	10
272	Geometría Analítica II	Gabriel Gutiérrez García	10
273	Geometría Analítica II	Fernando Brambila Paz	11
274	Geometría Analítica II	Eugenio Garnica Vigil	13
275	Geometría Analítica II	Pablo Barrera Sánchez	13
276	Geometría Analítica II	Elsa Puente Vázquez	15
277	Geometría Analítica II	Rolando Gómez Macedo	15
278	Geometría Analítica II	María de Lourdes Hernández Campos	15
279	Geometría Analítica II	Lorena Armas Sanabria	16
280	Geometría Analítica II	Herminio Suárez Quiroz	18

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
281	Geometría Analítica II	Camilo Camhaji García	20
282	Geometría Diferencial I	Pierre Michel Bayard	7
283	Geometría Diferencial I	Adriana Ortiz Rodríguez	9
284	Geometría Diferencial I	Federico Sánchez Bringas	14
285	Geometría Diferencial I	Esteban Librado Hernández Escamilla	19
286	Geometría Moderna I	Alejandro Bravo Mojica	9
287	Geometría Moderna I	Verónica Martínez de la Vega y Mansilla	10
288	Geometría Moderna I	Esteban Rubén Hurtado Cruz	10
289	Geometría Moderna I	José Antonio Gómez Ortega	12
290	Geometría Moderna I	Ramón Reyes Carrión	16
291	Geometría Moderna I	Saúl Arce Rocha	19
292	Geometría Moderna II	María Guadalupe Lucio Gómez-Maqueo	9
293	Geometría Moderna II	Isabel Alicia Hubard Escalera	15
294	Inferencia Estadística	Jimmy Hernández Morales	7
295	Inferencia Estadística	Miguel Ángel Chong Rodríguez	9
296	Inferencia Estadística	Edna Gabriela López Estrada	10
297	Inferencia Estadística	Carlos Erwin Rodríguez Hernández-Vela	10
298	Inferencia Estadística	Miguel Arturo Ballesteros Montero	10
299	Inferencia Estadística	Oscar Fontanelli Espinosa	11
300	Inferencia Estadística	Jesica Hernández Rojano	16
301	Inferencia Estadística	Lizbeth Román Padilla	16
302	Inferencia Estadística	Graciela Martínez Sánchez	17
303	Inferencia Estadística	César Almenara Martínez	19
304	Inglés I	Lilian Moreno Roldán	7
305	Inglés I	Diana Vianey Vargas Nieto	9
306	Inglés I	Silvia Loera Rivera	13
307	Inglés II	Lidia Fabiola Quevedo Rojas	8
308	Inglés V	María Teresa Pozos Delgado	7
309	Inglés V	Natalia Dan	11
310	Inglés VI	Alejandro Pérez Meléndez	10
311	Introducción a Ciencias de la Computación	Amparo López Gaona	8
312	Introducción a Ciencias de la Computación	Pedro Ulises Cervantes González	10
313	Introducción a Ciencias de la Computación	Francisco Valdés Souto	10
314	Introducción a Ciencias de la Computación	Canek Peláez Valdés	13

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
315	Introducción a Ciencias de la Computación	Luis Alberto Ramírez Bermudez	17
316	Introducción a Ciencias de la Computación	Virginia Teodosio Procopio	20
317	Introducción a la Física Cuántica	María de los Ángeles Ortiz Flores	10
318	Introducción a la Física Cuántica	Raúl Wayne Gómez González	10
319	Introducción a la Física Cuántica	Héctor Jesús Díaz Jiménez	10
320	Introducción a la Física Cuántica	Ignacio Campos Flores	14
321	Introducción a las Matemáticas Discretas	María del Pilar Valencia Saravia	7
322	Introducción a las Matemáticas Discretas	Rita Esther Zuazua Vega	8
323	Introducción a las Matemáticas Discretas	Andrés Carnero Bravo	9
324	Introducción a las Matemáticas Discretas	Mucuy-Kak del Carmen Guevara Aguirre	11
325	Introducción a las Matemáticas Discretas	María del Rocío Sánchez López	12
326	Introducción a las Matemáticas Discretas	Loiret Alejandría Dosal Trujillo	14
327	Introducción a las Matemáticas Discretas	Germán Benítez Bobadilla	15
328	Introducción a las Matemáticas Discretas	Gerardo Miguel Tecpa Galván	16
329	Introducción a las Matemáticas Discretas	Edgar Migueles Pérez	20
330	Investigación de Operaciones	Leonardo López Monroy	7
331	Investigación de Operaciones	María del Carmen Hernández Ayuso	9
332	Investigación de Operaciones	Ana Lilia Anaya Muñoz	10
333	Investigación de Operaciones	David Chaffrey Moreno Fernández	11
334	Investigación de Operaciones	Edgar Gil Hernández Díaz	16
335	Investigación de Operaciones	Hérica Sánchez Larios	18
336	Lógica Matemática I	Mariana Martínez González	8
337	Lógica Matemática I	Luis Jesús Turcio Cuevas	11
338	Lógica Matemática I	Cecilia Chávez Aguilera	14
339	Lógica Matemática II	José Gabriel Ocampo Márquez	10
340	Lógica Matemática II	María de Lourdes Guerrero Zarco	17
341	Manejo de Datos	Sonia Josefina Valery Lobo	7

La tabla continúa en la siguiente página

	Materia	**Profesor**	**Horario**
342	Manejo de Datos	José Carlos León Pérez	7
343	Manejo de Datos	Cristina Gónzalez Quintero	8
344	Manejo de Datos	Araceli Eugenia Mercado Fernández	9
345	Manejo de Datos	Miguel Ehécatl Morales Trujillo	10
346	Manejo de Datos	Manuel Alcántara Juárez	11
347	Manejo de Datos	Gerardo Avilés Rosas	12
348	Manejo de Datos	Miriam Sosa Díaz	13
349	Manejo de Datos	Virginia Teodosio Procopio	14
350	Manejo de Datos	Miguel Ángel Pérez León	17
351	Manejo de Datos	Miguel Manterola Obregón	18
352	Manejo de Datos	Silvestre Ruíz Salinas	19
353	Matemáticas Actuariales del Seguro de Personas I	Elisa Sugey Hérnandez Castañeda	7
354	Matemáticas Actuariales del Seguro de Personas I	José Enrique Pérez Salvador	20
355	Matemáticas Actuariales del Seguro de Personas I	Karina Vargas Cruz	21
356	Matemáticas Actuariales del Seguro de Personas II	Pedro Aguilar Beltrán	7
357	Matemáticas Actuariales del Seguro de Personas II	Alfonso Parrao Guzman	8
358	Matemáticas Actuariales del Seguro de Personas II	Oscar Aranda Martínez	8
359	Matemáticas Actuariales para Seguro de Daños, Fianzas y Reaseguro	José Luis López Escorcia	7
360	Matemáticas Actuariales para Seguro de Daños, Fianzas y Reaseguro	Jorge Luis Reyes García	7
361	Matemáticas Actuariales para Seguro de Daños, Fianzas y Reaseguro	Felipe Zamora Ramos	7
362	Matemáticas Actuariales para Seguro de Daños, Fianzas y Reaseguro	Elisa Sugey Hérnandez Castañeda	8
363	Matemáticas Actuariales para Seguro de Daños, Fianzas y Reaseguro	Irma Rocío Villa Valles	10
364	Matemáticas Actuariales para Seguro de Daños, Fianzas y Reaseguro	Juan Carlos Vargas Aguilar	19
365	Matemáticas Actuariales para Seguro de Daños, Fianzas y Reaseguro	Miguel Ángel Torres Ramírez	20
366	Matemáticas Financieras	Juan Francisco Carmona Sánchez	7
367	Matemáticas Financieras	Alejandro Carrillo Nolazco	8
368	Matemáticas Financieras	José Roberto de Jesús Gónzalez	17

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
369	Matemáticas Financieras	Jesús Alberto Rodríguez Sánchez	19
370	Matemáticas para las Ciencias Aplicadas I	Lina Elisa Santillán Espinoza	8
371	Matemáticas para las Ciencias Aplicadas I	Luis Enrique Quintanar Cortés	8
372	Matemáticas para las Ciencias Aplicadas I	Jesús Enrique Hernández Zavaleta	8
373	Matemáticas para las Ciencias Aplicadas I	Rodrigo Robles Montero	8
374	Matemáticas para las Ciencias Aplicadas I	Oscar García Zarco	10
375	Matemáticas para las Ciencias Aplicadas I	Omar Recillas Ayala	10
376	Matemáticas para las Ciencias Aplicadas I	Alejandro Villarreal López	10
377	Matemáticas para las Ciencias Aplicadas I	José Luis Gutiérrez Sánchez	12
378	Matemáticas para las Ciencias Aplicadas I	Soledad Ruiz Matus	17
379	Matemáticas para las Ciencias Aplicadas I	Raúl Jaime Torres Rojas	17
380	Matemáticas para las Ciencias Aplicadas II	Jorge Clouthier López	8
381	Matemáticas para las Ciencias Aplicadas II	Jonathan Galván Colín	8
382	Matemáticas para las Ciencias Aplicadas II	Manuel Velasco Juan	10
383	Matemáticas para las Ciencias Aplicadas II	Oscar Olicón Hernández	10
384	Matemáticas para las Ciencias Aplicadas II	Arturo Velasco Pelayo	10
385	Matemáticas para las Ciencias Aplicadas II	Juan Carlos Balleza García	10
386	Matemáticas para las Ciencias Aplicadas II	Juan Carlos Díaz Patiño	11
387	Matemáticas para las Ciencias Aplicadas II	Pedro Eduardo Miramontes Vidal	14
388	Matemáticas para las Ciencias Aplicadas II	Ricardo Zavaleta Madrid	16
389	Matemáticas para las Ciencias Aplicadas II	Raúl Jaime Torres Rojas	16
390	Matemáticas para las Ciencias Aplicadas II	Zeus Alberto Valtierra Quintal	17

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
391	Matemáticas para las Ciencias Aplicadas III	Luciano Martínez Balbuena	12
392	Matemáticas para las Ciencias Aplicadas III	Daniel Castillo Rodríguez Arana	18
393	Matemáticas para las Ciencias Aplicadas IV	Jesús Edel Cereceres Delgado	8
394	Matemáticas para las Ciencias Aplicadas IV	Mariana Soledad Centeno Sierra	10
395	Matemáticas para las Ciencias Aplicadas IV	Isaac Arelio Ríos	10
396	Matemáticas para las Ciencias Aplicadas IV	Temilotzin Ibarra Delgadillo	10
397	Matemáticas para las Ciencias Aplicadas IV	Faustino Sánchez Garduño	12
398	Matemáticas para las Ciencias Aplicadas IV	Jorge Avella Martínez	16
399	Mecánica Analítica	Jerónimo Alonso Cortez Quezada	8
400	Mecánica Analítica	Rosa María Méndez Vargas	10
401	Mecánica Analítica	María Luisa Marquina Fábrega	10
402	Mecánica Analítica	Graciela Velasco Herrera	17
403	Mecánica Analítica	Lorea Chaos Cador	18
404	Mecánica Cuántica	Luis Fernando Magaña Solís	8
405	Mecánica Cuántica	Héctor Hernández Coronado	10
406	Mecánica Cuántica	Fernando Matías Moreno Yntriago	10
407	Mecánica Cuántica	Enriqueta Hernández Saldaña	10
408	Mecánica Cuántica	David García Gudiño	16
409	Mecánica Cuántica	Enrique Yépez Mulia	18
410	Mecánica Vectorial	Jorge Ramón Soto Mercado	7
411	Mecánica Vectorial	Isaías Rodríguez Aguirre	7
412	Mecánica Vectorial	Roberto Alejandro Ruelas Mayorga	7
413	Mecánica Vectorial	Manuel Gerardo Quintana García	8
414	Mecánica Vectorial	Consuelo García Alcántara	10
415	Mecánica Vectorial	Alejandro Valderrama Zaldívar	10
416	Mecánica Vectorial	Carlos Málaga Iguiñiz	10
417	Mecánica Vectorial	Darío Núñez Zúñiga	10
418	Mecánica Vectorial	Juan Manuel Eugenio Ramírez de Arellano Niño-Rincón	10
419	Mecánica Vectorial	Vicenta Sánchez Morales	10

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
420	Mecánica Vectorial	Susana Orozco Segovia	10
421	Mecánica Vectorial	Erick Leonardo Patiño Jaidar	14
422	Mecánica Vectorial	Francisco Nettel Rueda	18
423	Mecánica Vectorial	David García Gudiño	18
424	Mercados Financieros y Valuación de Instrumentos	Leticia Guadalupe Cienfuegos Blancas	7
425	Mercados Financieros y Valuación de Instrumentos	Jorge Luis Saviñon Basurto	8
426	Mercados Financieros y Valuación de Instrumentos	Guillermo Rolando Rosas Figueroa	8
427	Mercados Financieros y Valuación de Instrumentos	Eduardo Selim Martínez Mayorga	9
428	Mercados Financieros y Valuación de Instrumentos	María del Rosario Espinosa Tufiño	16
429	Mercados Financieros y Valuación de Instrumentos	Yurguen Hugo Camargo Serafin	19
430	Métodos Cuantitativos en Finanzas	Nayeli Castillo Carranco	7
431	Métodos Cuantitativos en Finanzas	Jorge Luis Silva Haro	7
432	Métodos Cuantitativos en Finanzas	Ramiro Alberto Domínguez Ameneyro	7
433	Métodos Cuantitativos en Finanzas	Daniela Alejandra González Ramírez	8
434	Métodos Cuantitativos en Finanzas	Jorge Humberto Del Castillo Spíndola	15
435	Métodos Cuantitativos en Finanzas	Fernando Díaz López	19
436	Modelado y Programación	José de Jesús Galaviz Casas	10
437	Modelado y Programación	Gustavo Arturo Márquez Flores	14
438	Modelado y Programación	Cinthia Rodríguez Maya	18
439	Modelos de Supervivencia y de Series de Tiempo	Gerardo Sisniega Lira	7
440	Modelos de Supervivencia y de Series de Tiempo	David Chaffrey Moreno Fernández	8
441	Modelos de Supervivencia y de Series de Tiempo	Claudia González González	10
442	Modelos de Supervivencia y de Series de Tiempo	Margarita Elvira Chávez Cano	10
443	Modelos de Supervivencia y de Series de Tiempo	Guillermina Eslava Gómez	16
444	Modelos de Supervivencia y de Series de Tiempo	Rubén Ugalde Franco	17
445	Modelos no Paramétricos y de Regresión	Margarita Elvira Chávez Cano	9

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
446	Modelos no Paramétricos y de Regresión	Javier Santibañez Cortés	10
447	Modelos no Paramétricos y de Regresión	Jose Salvador Zamora Muñoz	10
448	Modelos no Paramétricos y de Regresión	Lizbeth Naranjo Albarrán	11
449	Modelos no Paramétricos y de Regresión	Jaime Vázquez Alamilla	12
450	Modelos no Paramétricos y de Regresión	José Antonio Flores Díaz	12
451	Modelos no Paramétricos y de Regresión	Fernando Daniel Pérez Arriaga	17
452	Modelos no Paramétricos y de Regresión	Jimmy Hernández Morales	18
453	Modelos no Paramétricos y de Regresión	Rosa Daniela Chávez Aguilar	18
454	Modelos no Paramétricos y de Regresión	Alejandro Martínez López	19
455	Muestreo	José Antonio Flores Díaz	11
456	Óptica	Andrea Luisa Aburto Espina	8
457	Óptica	José Ignacio Jiménez Mier y Terán	8
458	Óptica	Víctor Hugo Meza Laguna	10
459	Óptica	Nadia Ramírez Cruz	10
460	Óptica	José Luis Pérez Mazariego	10
461	Óptica	Enrique López Moreno	12
462	Óptica	Manuel Campos García	18
463	Pensiones Privadas	Carlos Fernando Lozano Nathal	7
464	Pensiones Privadas	José Antonio Valencia Trujillo	7
465	Pensiones Privadas	Ernesto Rosas Garcia	8
466	Pensiones Privadas	Jesús Paz Méndez	18
467	Planeación Estratégica	Carlos Orozco Rocha	7
468	Planeación Estratégica	Mónica Iliana Sánchez Zaragoza	7
469	Probabilidad I	Jesús Alberto Rodríguez Sánchez	7
470	Probabilidad I	Jaime Vázquez Alamilla	8
471	Probabilidad I	María Asunción Begoña Fernández Fernández	9
472	Probabilidad I	Yuri Salazar Flores	9
473	Probabilidad I	Francisco Pérez Carbajal	10
474	Probabilidad I	Perla Marysol Ruiz Arias	10

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
475	Probabilidad I	María del Pilar Alonso Reyes	10
476	Probabilidad I	Bibiana Obregón Quintana	14
477	Probabilidad I	Federico Lasa Gonsebatt	16
478	Probabilidad I	Freddy Palma Mancilla	17
479	Probabilidad I	Juan González Hernández	19
480	Probabilidad II	Sandra Palau Calderón	8
481	Probabilidad II	Alberto Contreras Cristán	10
482	Probabilidad II	Marco Arieli Herrera Valdez	10
483	Probabilidad II	Luis Antonio Rincón Solís	10
484	Probabilidad II	Marco Arieli Herrera Valdez	13
485	Probabilidad II	Iván Ixcoatl Juárez López	14
486	Probabilidad II	Rafael Miranda Cordero	16
487	Probabilidad II	Miguel Ángel García Álvarez	17
488	Probabilidad II	Federico Lasa Gonsebatt	18
489	Probabilidad II	Francisco Daniel Ramírez Calixto	19
490	Probabilidad II	Sergio Iván López Ortega	19
491	Procesos Estocásticos I	Fernando Guerrero Poblete	8
492	Procesos Estocásticos I	Guadalupe Carrasco Licea	9
493	Procesos Estocásticos I	María Clara Fittipaldi	10
494	Procesos Estocásticos I	Rodrigo Quijón Hipólito	10
495	Procesos Estocásticos I	Adrián González Casanova Soberon	10
496	Procesos Estocásticos I	Sergio Iván López Ortega	15
497	Procesos Estocásticos I	Daniel Antonio Márquez Vázquez	19
498	Procesos Estocásticos I	Jose Rodrigo Iniesta Miranda	19
499	Productos Financieros Derivados	Guillermo Rolando Rosas Figueroa	7
500	Productos Financieros Derivados	Jesús Agustín Cano Garcés	10
501	Programación	Víctor Manuel Carreón Calderón	7
502	Programación	Jessica Santizo Galicia	8
503	Programación	Mario González Ruiz	8
504	Programación	Armando Orozco Cortés	8
505	Programación	Gustavo Adolfo García Cano	8
506	Programación	Sonia Josefina Valery Lobo	9
507	Programación	Luis Enrique Serrano Gutiérrez	10
508	Programación	Jaime Ayala Pérez	11
509	Programación	Gerardo Avilés Rosas	13
510	Programación	José Alfredo Cobián Campos	13
511	Programación	Antonio Carrillo Ledesma	13
512	Programación	Ricardo Castañeda Martínez	14

La tabla continúa en la siguiente página

	Materia	**Profesor**	**Horario**
513	Programación	Alejandro Felipe Zarate Pérez	17
514	Programación	Alejandro Carrillo Nolazco	17
515	Redes de Computadoras	Luis Enrique Serrano Gutiérrez	8
516	Redes de Computadoras	María Teresa Canseco Rodríguez	9
517	Relatividad	Mirna Villavicencio Torres	8
518	Relatividad	Saúl Noé Ramos Sánchez	8
519	Relatividad	Eduardo Peinado Rodríguez	8
520	Relatividad	Juan Manuel García Islas	10
521	Relatividad	Yuri Bonder Grimberg	10
522	Relatividad	Ismael Oviedo De Julian	10
523	Relatividad	Mariano Chernicoff Minsberg	14
524	Relatividad	Francisco Nettel Rueda	16
525	Relatividad	Sebastián Nájera Valencia	16
526	Relatividad	Pedro Antonio Sánchez Serrano	18
527	Relatividad	Gustavo Enrique García de Jesús	20
528	Seguridad Social	Silvia Leticia Malpica Flores	7
529	Seguridad Social	Mariana Velasco Rodriguez	8
530	Seminario de Álgebra A	Juan Orendain Almada	8
531	Seminario de Álgebra A	Emilio Esteban Lluis Puebla	11
532	Seminario de Análisis Matemático A	Mónica Alicia Clapp Jiménez-Labora	10
533	Seminario de Aplicaciones Actuariales	Carlos Contreras Cruz	7
534	Seminario de Aplicaciones Actuariales	Víctor Manuel Solís Nájera	7
535	Seminario de Aplicaciones Actuariales	Eduardo Torres Luna	7
536	Seminario de Aplicaciones Actuariales	Mónica Iliana Sánchez Zaragoza	8
537	Seminario de Aplicaciones Actuariales	Carlos Díaz Ávalos	10
538	Seminario de Aplicaciones Actuariales	Ricardo Ramírez Aldana	10
539	Seminario de Aplicaciones Actuariales	Irma Rocío Villa Valles	11
540	Seminario de Aplicaciones Actuariales	Omar de la Riva Torres	18
541	Seminario de Ciencias de la Computación	Jonathan Banfi Vázquez	8
542	Seminario de Ciencias de la Computación	José David Flores Peñaloza	10
543	Seminario de Ciencias de la Computación	Jorge Luis Ortega Arjona	10
544	Seminario de Ciencias de la Computación	Salvador López Mendoza	10

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
545	Seminario de Ciencias de la Computación	Sergio Rajsbaum Gorodezky	11
546	Seminario de Ciencias de la Computación	Jorge Luis Ortega Arjona	12
547	Seminario de Ciencias de la Computación	Luis Alberto Ramírez Bermudez	15
548	Seminario de Estadística I	Alonso Baranda Lozada	8
549	Seminario de Estadística I	Rafael Miranda Cordero	10
550	Seminario de Estadística I	Ruth Selene Fuentes García	10
551	Seminario de Estadística I	Guillermina Eslava Gómez	13
552	Seminario de Geometría A	Pablo Rosell González	9
553	Seminario de Geometría A	Saúl Juárez Mena	13
554	Seminario de Geometría A	Augusto Cabrera Becerril	18
555	Seminario de Topología A	Alejandro Illanes Mejía	9
556	Seminario de Topología A	Enrique Guillermo Bazúa Durán	14
557	Seminario de Topología A	Luis Manuel Venegas Grajales	18
558	Seminario Matemáticas Aplicadas I	Adrián González Casanova Soberon	8
559	Seminario Matemáticas Aplicadas I	Hortensia Galeana Sánchez	10
560	Seminario Matemáticas Aplicadas I	Pedro José Sobrevilla Moreno	18
561	Seminario sobre Enseñanza de las Matemáticas I	Carmen Martínez-Adame Isaís	9
562	Seminario sobre Enseñanza de las Matemáticas I	Francisco de Jesús Struck Chávez	10
563	Seminario sobre Enseñanza de las Matemáticas I	José Rafael Martínez Enríquez	10
564	Seminario sobre Enseñanza de las Matemáticas I	Vicente Carrión Velázquez	19
565	Seminario Sobre Enseñanza Matemáticas III	Alejandro Ricardo Garciadiego Dantan	9
566	Taller de Herramientas Computacionales	María Fernanda Sánchez Puig	8
567	Temas Selectos de Economía	Luis Alberto Quezada Téllez	8
568	Teoría de Gráficas	Jesús Alva Samos	8
569	Teoría de Gráficas	Mario Aurelio Yañez Molina	19
570	Teoría de los Conjuntos I	David Meza Alcántara	8
571	Teoría de los Conjuntos I	Mauricio Salinas Rodríguez	9
572	Teoría de los Conjuntos I	Ángel Tamariz Mascarúa	10
573	Teoría de los Conjuntos I	Mario Francisco Rosales González	14
574	Teoría de los Conjuntos I	José Adrián Gallardo Quiróz	15

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
575	Teoría de los Números I	Alejandro De Las Peñas Castaño	8
576	Teoría de los Números I	Elhoim Llorente Sumano y Ramírez	17
577	Teoría del Riesgo	Luis Alberto Olvera García	7
578	Teoría del Riesgo	Daniel Cervantes Filoteo	8
579	Teoría del Riesgo	Juan Diego Amaya Figueroa	8
580	Teoría del Riesgo	Alejandro Santoyo Cano	8
581	Teoría del Riesgo	David Josafat Santana Cobian	10
582	Teoría del Riesgo	Yuri Salazar Flores	10
583	Teoría del Riesgo	Jorge Luis Saviñon Basurto	21
584	Teoría del Seguro	Mauricio Gabriel Arredondo Fernández Cano	7
585	Teoría del Seguro	Karen Lanzguerrero Obeid	9
586	Termodinámica	Carlos Enrique Román Velázquez	8
587	Termodinámica	María Michel Duque Vega	8
588	Termodinámica	Adriano Valdés Gómez	18
589	Topología I	Gerardo Acosta García	8
590	Topología I	Sergio Macías Álvarez	9
591	Topología I	Grissel Santiago González	12
592	Topología I	Patricia Pellicer Covarrubias	14
593	Topología I	María Isabel Puga Espinosa	16
594	Topología I	Jesús Manuel Mayorquín García	20
595	Topología II	Jorge Marcos Martínez Montejano	12
596	Topología II	Héctor Jiménez Sánchez	17
597	Topología III	Mario Eudave Muñoz	11
598	Variable Compleja I	Emigdio Martínez Ojeda	10
599	Variable Compleja I	Ricardo Gómez Aiza	10
600	Variable Compleja I	Héctor Fidencio Sánchez Morgado	10
601	Variable Compleja I	Jefferson Edwin King Dávalos	11
602	Variable Compleja I	Oscar Alfredo Palmas Velasco	11
603	Variable Compleja I	Carisa Cano Figueroa	13
604	Variable Compleja I	Pavel Ramos Martínez	14
605	Variable Compleja I	Roberto Pichardo Mendoza	16
606	Variable Compleja I	Mico Djurdjevic	18

Tabla D.1: *Se muestra la matriz con la asignación final de los horarios. Cada renglón tiene la información de un grupo con una materia, profesor y horario asignado.*

Apéndice E

Observaciones / Notas

1. La matriz `mat_posibles_url` se define con un tamaño fijo antes de correr el algoritmo para que no se demore por tener un objeto que va cambiando de tamaño, por lo que al final de haberle aplicado la función se le deben de quitar los renglones que no tienen información.

2. La función *casos_alumnos* convierte los *NA* de la columna *Alumnos* de *m_grande* en ceros pero al generar *m_grande_total* y pasarla por la función *limpia_m_grande* se eliminan los *NA* y se cambian por ceros por lo tanto no es necesaria la función *casos_alumnos*, basta pasar la columna correspondiente a *Alumnos* de *m_grande_total*.

3. Cuando se hacen comparaciones se toman los valores reales y se les restan los valores simulados $(Reales - \mathbb{E}[Simulados])$

4. Con las gráficas *heatmap* se revisa si el modelo es adecuado o si se debe modificar algo. Se espera que las gráficas sean de color claro ya que nos interesa que el número de grupos y alumnos simulados se parezca al real.

5. Se tienen dos tipos de matrices las cuales llamaremos *m_objetivo* y *m_definición*; las matrices *m_objetivo* son las que tienen la información que se utiliza para la asignación; las matrices *m_definición* nos sirven para dos cosas:

   a) Respaldo de la descripción de cada columna

   b) Para guardar los índices en los que se encuentran las columnas

6. Las matrices tipo *m_definición* son:

   a) mat_def_columnas_MG

   b) mat_def_grupos_reales

   c) mat_def_grupos_simulados

7. Las matrices tipo *m_objetivo* son:

   a) m_grande

   b) m_grande_total

c) ...

8. La función *checa_ind_materia* se encarga de obtener lo índices de las columnas de las matrices tipo *m_definición* para poder sacar información de *m_grande* o de *m_grande_total*.

9. Para las simulaciones se utiliza la información anterior a la del semestre que se quiere simular para no tener información real dentro de los datos para la simulación.

10. En caso de querer elegir la capacidad del salón se va a elegir la mayor de sus capacidades (comparando las capacidades que se han tenido a lo largo de varios semestres).

11. Las matrices *m_grande* y de *m_grande_total* tienen información real.

12. En los ciclos que recorren renglones y columnas de matrices, siempre es más rápido hacer (de afuera hacia adentro) primero las columnas y luego los renglones.

    Si se tiene una matriz con entradas (i, j) entonces:

```
for(j){
 for(i){
    m[i,j]
  }
}
```

Código E.1: *Ejemplo de ciclo for*

13. El vector *vec_nom_materias_total* tiene los nombres de las materias, sin repeticiones, que se utiliza para las simulaciones.

14. El vector *vec_excepciones* tiene las posibles ecxepciones en las que las funciones que extraen información pueden caer, de esta manera se pueden generar nuevas funciones para corregir esos casos.

15. La siguiente imagen es el resultado de la función *imprime_info_idiomas* la cual muestra la información de los idiomas. Dicha función arroja un vector con los semestres que requieren modificación.

```
La matriz m_grande del semestre  20081  no tiene clases de  Inglés
La matriz m_grande del semestre  20082  no tiene clases de  Inglés
La matriz m_grande del semestre  20091  no tiene clases de  Inglés
La matriz m_grande del semestre  20092  no tiene clases de  Inglés
La matriz m_grande del semestre  20101  no tiene clases de  Inglés
La matriz m_grande del semestre  20102  no tiene clases de  Inglés
La matriz m_grande del semestre  20111  no tiene clases de  Inglés
La matriz m_grande del semestre  20112  no tiene clases de  Inglés
La matriz m_grande del semestre  20121  no tiene clases repetidas de  Inglés
La matriz m_grande del semestre  20122  no tiene clases repetidas de  Inglés
La matriz m_grande del semestre  20131  no tiene clases repetidas de  Inglés
La matriz m_grande del semestre  20132  no tiene clases repetidas de  Inglés
La matriz m_grande del semestre  20141  no tiene clases repetidas de  Inglés
La matriz m_grande del semestre  20142  no tiene clases repetidas de  Inglés
La matriz m_grande del semestre  20151  no tiene clases repetidas de  Inglés
En el semestre  20152  se tienen  2  clases repetidas de  Inglés
En el semestre  20161  se tienen  3  clases repetidas de  Inglés
En el semestre  20162  se tienen  4  clases repetidas de  Inglés
En el semestre  20171  se tienen  1  clases repetidas de  Inglés
En el semestre  20172  se tienen  2  clases repetidas de  Inglés
En el semestre  20181  se tienen  4  clases repetidas de  Inglés
En el semestre  20182  se tienen  3  clases repetidas de  Inglés
La matriz m_grande del semestre  20191  no tiene clases repetidas de  Inglés
En el semestre  20192  se tienen  1  clases repetidas de  Inglés
En el semestre  20201  se tienen  1  clases repetidas de  Inglés
```

Figura E.1: *Resumen de clases de ingles antes de modificación*

Con esta información se decidió observar caso por caso los renglones que requieren modificación para la matriz *m_grande*

16. Debido a la situación en la que estamos viviendo actualmente, ahora más que nunca es necesario tener un programa para la asignación de horarios que permita la realización de las asignaciones sin tener la necesidad de hacer reuniones en persona, ya que al proseguir con las medidas de distanciamiento social, las reuniones antiguamente hechas en persona se tendrían que hacer por medio de alguna plataforma digital las cuales no necesariamente son las más óptimas ya que dependen de la señal de todos los participantes para que haya una comunicación de manera fluída. Debido a ésto, el programa es una buena solución.

17. Al hacer las simulaciones del número de alumnos el redondeo es hacia arriba, usando la función *ceiling*.

18. El vector *vec_nom_materias_total*, que contiene el nombre de las materias se definió en la lista *param* para poder tomarlo en las diferentes funciones.

19. Para resolver un problema, pensar en los pasos en los que se puede dividir dicho problema, usualmente se requieren entre 3 y 8 pasos o casos para obtener un producto final. Para cada paso hacer una función.

    Se tienen dos posibles estructuras:

    a) La función del paso n manda a llamar a la del paso $n-1$.

       Ej.

       simula_grupos{*simula_gpos_1_sem* {*simula_gpos_1_materia* {*simula_tam_gpo*

    b) Se tiene una función principal que manda a llamar a las funciones de cada paso:

       Ej.

```
gen_asignacion_completa <- function(sem_ini,sem_fin
  ){
  # Se carga y se limpia la lista de urls (para no
     tener paginas sin informacion,...)
  list_url <- Actualiza_list_url(list_url)

  # Se obtiene "m_grande" y se genera un archivo
     para cada semestre
  for(k in 1:length(semestres)){
    sem_info <- semestres[k]
    directorio_info[k] <- gen_m_grande(sem_info,
       list_url)
  }

  # Se genera el esqueleto del semestre que se
     quiere obtener
```

```
        mat_esqueleto <- gen_esqueleto(directorio_info,
          param)

        # Se genera la matriz de solicitudes de todos los
          profesores
        mat_solicitudes <- gen_solicitudes(param)

        # Se genera la matriz de asignaciones de todos
          los profesores
        mat_asignaciones <- gen_asignacion(mat_esqueleto,
          mat_solicitudes,param)

        return(mat_asignaciones)
    }
}
```

Código E.2: *Ejemplo de estructura de funciones*

20. Pudiera ser que haya un apéndice con "Observaciones" utlizando las notas escritas.

21. Todo lo que se escriba debe tener un propósito, sino quitarlo.

22. La información que se puede encontrar actualmente (debido a la pandemia) en las páginas web de los horarios de la Facultad no es la misma que la mostrada a lo largo del trabajo ya que ahora no se tiene información del salón, o del número de alumnos inscritos por materia, ni los lugares disponibles por grupo.

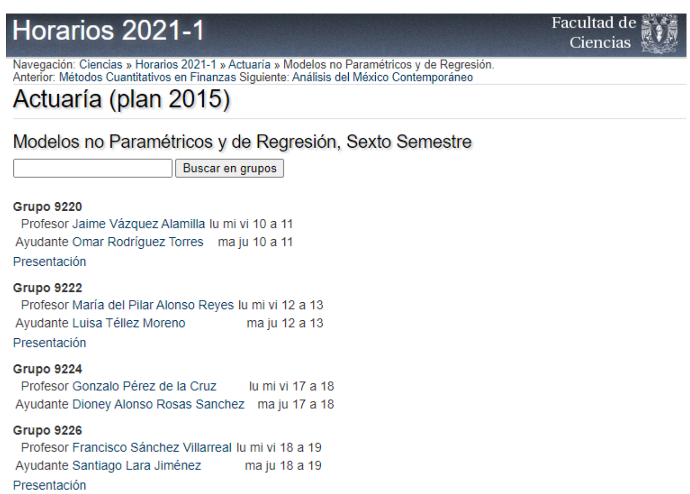
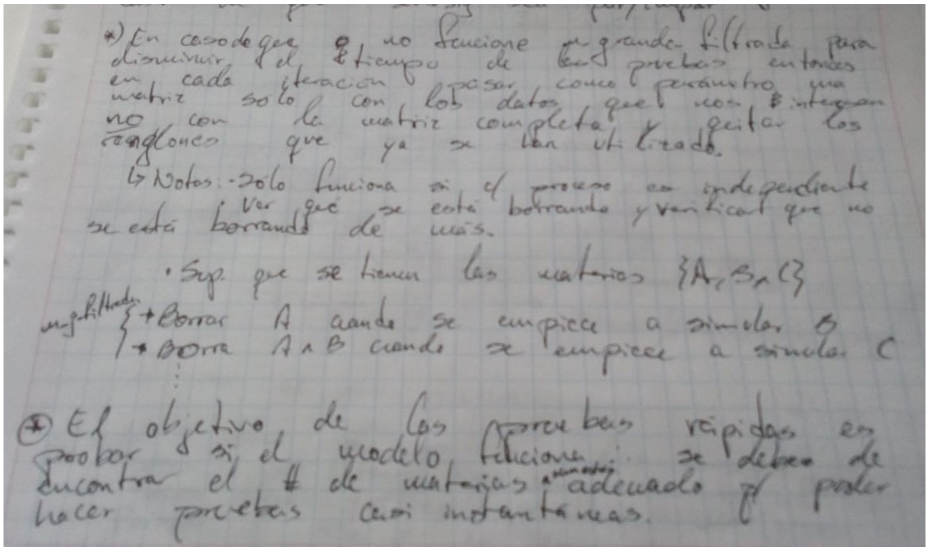


Figura E.2: *Ejemplo de horarios de semestre 2021-1*

23. Notas de T26


*) En caso de que ningún filtrado para disminuir el tiempo de las pruebas entonces en cada iteración pasar como parámetro una matriz sólo con los datos que se integran no con la matriz completa y quitar los renglones que ya se han utilizado.

↳ Notas: - Sólo funciona si el proceso es independiente
· Ver que se está borrando y verificar que no se está borrando de más.

· Sup que se tienen las matrices {A, B, C}

ningún filtrado { + Borrar A cuando se empiece a simular B
+ Borra A ∧ B cuando se empiece a simular C

⊛ El objetivo de las pruebas rápidas es probar si el modelo funciona. Se debe de encontrar el # de matrices adecuado p/ poder hacer pruebas casi instantáneas.

Figura E.3: *Notas de T26*

24. En caso de tener subsecciones: entre 3 y 4

25. Ser muy directa al escribir, pero explicar mucho más (platicar más). No hacer enunciados tan largos. No puede haber párrafos formados por un sólo enunciado. Escribir una idea por enunciado. No sólo escribir en párrafos, utilizar listas, tablas, ...

26. La estructura de cada párrafo debe ser de tipo *reloj de arena*. Ir de lo general a lo particular y volver a lo general con una conclusión.

27. Un enunciado equivale a una idea. Un párrafo equivale a un conjunto de ideas comunes.

28. Sea $D = \frac{r-s}{s}$, donde r son datos reales, s datos simulados y D la diferencia relativa, se busca que $D \in \left[-\frac{1}{2}, \frac{1}{2}\right]$.

29. Ejemplo del uso del comando *Roxygen* para comentar las funciones en *R*.

```
#' Add together two numbers
#'
#' @param x A number
#' @param y A number
#' @return The sum of \code{x} and \code{y}
#' @examples
#' add(1, 1)
#' add(10, 1)
add <- function(x, y) {
  x + y
}
```

Figura E.4: *Ejemplo de Roxygen*

30. Escribir en el archivo de LaTeX pequeños comentarios de la idea que se quiere transmitir en cada párrafo (de 2 a 3 palabras claves). Ésto sirve para referencias futuras y

para ordenar los párrafos con mayor facilidad.

31. Escribir párrafos de 2 a 3 enunciados completos, no dejar enunciados solos a menos que contengan información muy importante.
32. En caso de tener más de 10 referencias bibliográficas utilizar *Mendeley* para genera un archivo *.bib* y ponerlo en la tesis para tener la bibliografía.
33. Cuidar el tamaño de letra en las gráficas que se pongan
34. No poner abreviaturas en los títulos.
35. La imagen 3.1 tiene título en inglés, se tienen 2 opciones: dejarlo así o buscar cómo cambiarlo.
36. Recordar la diferencia entre:
    - Número de alumnos inscritos
    - Número de alumnos reales
    - Número de alumnos que toman clase por cada horario (no se toman en cuenta los alumnos que empalman clases)
37. Para la elección de q_1 y q_2 se debe darle prioridad a la varianza no al mín y al máx porque se pueden tener casos en los que el mín y el máx estén muy cercanos a cero (gráfica superior) pero su varianza es grande. Queremos que la varianza se encuentre alrededor del cero (gráfica inferior).

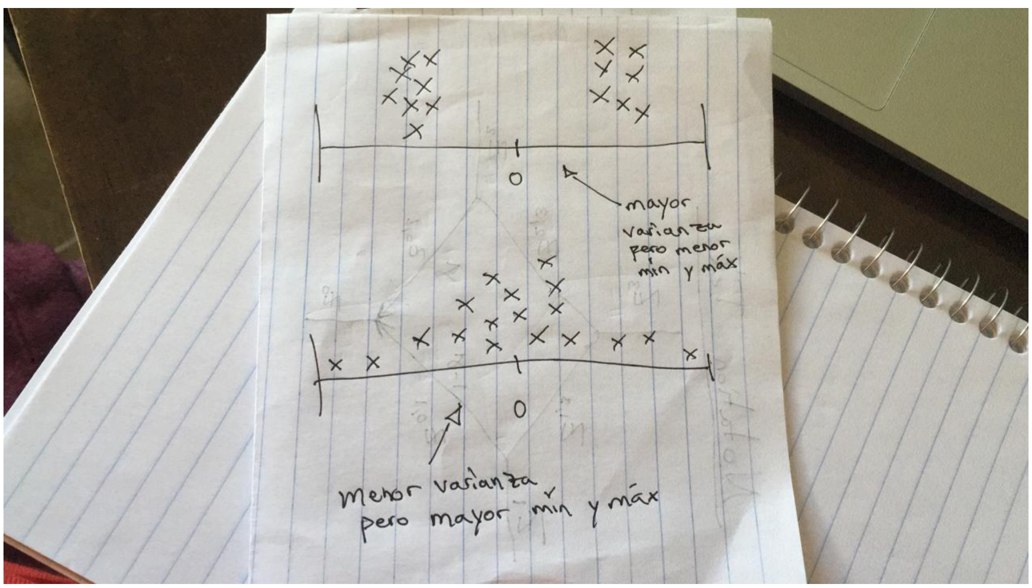


Figura E.5: *Ejemplo de varianza*

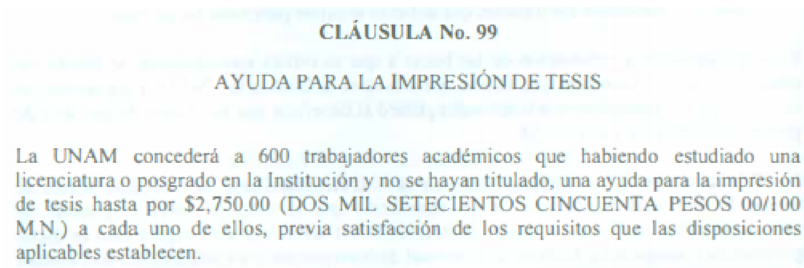
38. Preferir sacrificar el B/N en las imágenes impresas para tener una mejor versión digital a color.
39. Guardar figuras hechas en R con el comando: *dev.print(pdf, “Figures/Fig_Examples_of_GB_distributions height=5)*

40. Arrigo dijo que posiblemente alguien se va a quejar de no tomar en cuenta la preferencia de los profesores al realizar las solicitudes.

41. Un histograma nos muestra la representación de la distribución empírica de un conjunto de datos. Cada barra en el histograma representa la frecuencia de un intervalo sobre el rango de las observaciones que se tienen.

42. Cláusula 99 CCTPA: Ayuda para la impresión de la tesis.

    `https://www.personal.unam.mx/Docs/Contratos/AAPAUNAM20132015.pdf`



CLÁUSULA No. 99

AYUDA PARA LA IMPRESIÓN DE TESIS

La UNAM concederá a 600 trabajadores académicos que habiendo estudiado una licenciatura o posgrado en la Institución y no se hayan titulado, una ayuda para la impresión de tesis hasta por $2,750.00 (DOS MIL SETECIENTOS CINCUENTA PESOS 00/100 M.N.) a cada uno de ellos, previa satisfacción de los requisitos que las disposiciones aplicables establecen.

Figura E.6: *Cláusula 99 CCTPA: Ayuda para la impresión de la tesis*

43. Equivalencias de nombres para estadística:

    a) Estadística I - Inferencia Estadística

    b) Estadística II - Modelos no Paramétricos y de Regresión

    c) Estadística III - Modelos de Supervivencia y de Series de Tiempo

44. La frecuencia relativa en los histogramas no refleja directamente el porcentaje. Se debe multiplicar el valor del eje Y por el ancho del intervalo por 100 para obtener cifras en porcentaje. El área total de las barras sumará 1 (17).

45. No confundir las carpetas de *Figuras* del GitHub con la del pdf.

46. Ya no son necesarias las pruebas de bondad de ajuste porque los tamaños de grupo se van a simular con respecto a los profesores. Ver $T_{32}xx$)

47. Los archivos *README* sirven para explicar las cosas a los demás.

48. Si los grupos pequeños dan muchos problemas podemos considerar quitarlos.

49. Las materias que se actualizaron o cambiaron de nombre se pueden ver en Capítulo A.

50. Arrigo dijo que posiblemente alguien se va a quejar del hecho de que actualmente las inscripciones ya no se hacen con tira de materias firmada.

51. El comando `\figurename{~\ref{nom_figura}}` imprime la palabra *Figura* antes del número correspondiente a la figura de la referencia.

52. El comando `\chaptername{~\ref{nom_capitulo}}` imprime la palabra *Capítulo* antes del número correspondiente al capítulo de la referencia.

53. El comando `\tablename{~\ref{nom_tabla}}` imprime la palabra *Tabla* antes del número correspondiente a la tabla de la referencia.

54. En los 3 comandos anteriores la ~ sirve para poner un espacio entre el nombre y el número.

55. Los comandos `\subsecname{\ref{nom_subseccion}}`, `\secname{\ref{nom_seccion}}`, `\subsectionname{\ref{nom_subseccion}}` y `\subsectionname{\ref{nom_seccion}}` no existen.

56. Para cada figura, al momento de explicarla, pensar en el mensaje principal que se quiere transmitir y *dejarla hablar* por sí sola.

57. Nos interesa más el comportamiento de semestres más recientes. Darle más peso a ellos en las figuras.

58. Utilizar la coma de Oxford en caso de confusión o si el último elemento es compuesto. Ej. Finanzas II, Procesos Estocásticos I, y Probabilidad y Estadística.

59. Imagen que muestra el uso de Plan de estudio con sus diferentes variantes *Plan de Estudio, Plan de Estudios, Planes de Estudios, Planes de Estudio*. Los links en donde se encuentran esos nombres son: `https://www.dgae-siae.unam.mx/educacion/planes.php` y `https://www.dgae.unam.mx/planes/licenciatura.html`.

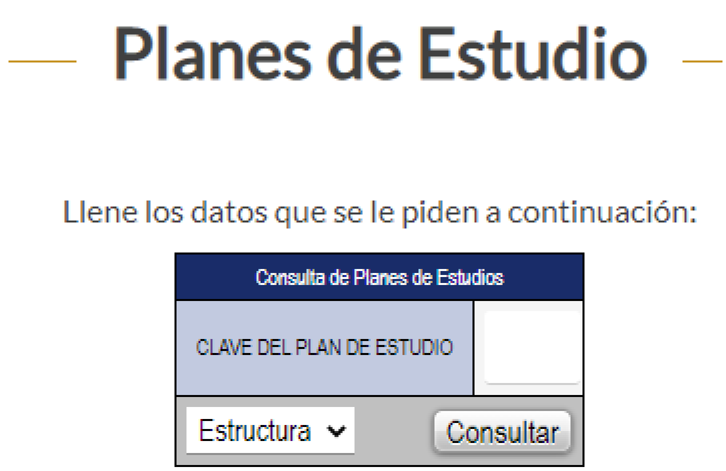
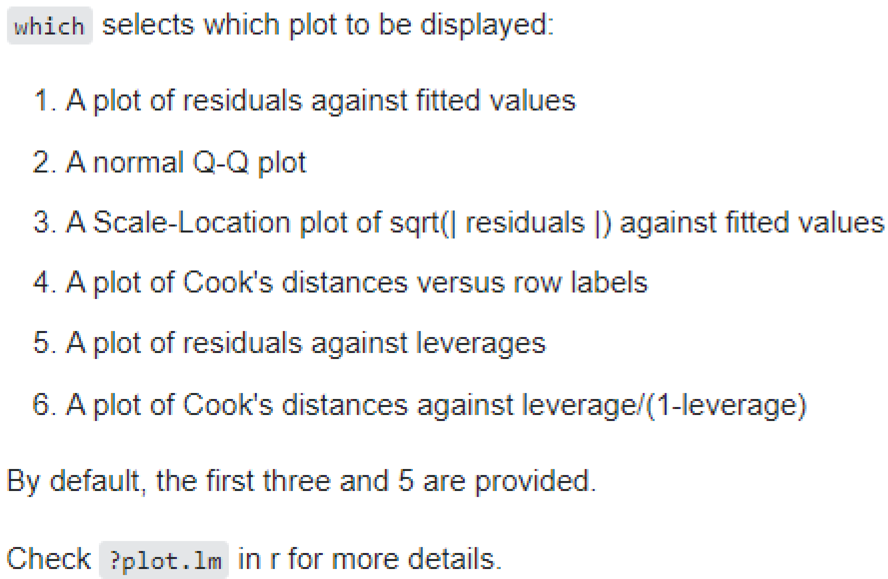


Figura E.7: *Nombres planes de estudio*

60. No usar palabras despectivas como: restante,...

61. En el modelo de mezcla de Normales tenemos el comando `plot(mixmdl,which = 2)`, la opción `which` se encarga de seleccionar el tipo de gráfica que se muestra. `https://stackoverflow.com/questions/29044055/plot-which-parameter-where-in-r`



`which` selects which plot to be displayed:

1. A plot of residuals against fitted values
2. A normal Q-Q plot
3. A Scale-Location plot of sqrt(| residuals |) against fitted values
4. A plot of Cook's distances versus row labels
5. A plot of residuals against leverages
6. A plot of Cook's distances against leverage/(1-leverage)

By default, the first three and 5 are provided.

Check `?plot.lm` in r for more details.

Figura E.8: *which in plot*

62. Uso de mayúsculas:

    - RAE: `https://www.rae.es/dpd/may%C3%BAsculas#33c`
    - Otro: `http://iesbinef.educa.aragon.es/lengua/ortografia/reglas/reglama.htm`

63. Uso de mayúsculas después de dos puntos RAE:

    `https://www.rae.es/dpd/dos%20puntos`

64. How to write your PhD thesis (without going insane) `https://www.youtube.com/watch?v=pM6orL-bGDc&ab_channel=JamesHaytonPhD`:

    - Definir tiempos de trabajo y tiempos de trabajo.
    - Ser constante. Escribir al menos una página al día.
    - Escribir más de las áreas en las que se tiene mayor conocimiento que en temas que no se conocen al 100 %.
    - Si se tiene un nivel de habilidad medio y el nivel del problema/reto es alto, entonces basta que uno se concentre en el problema para poder resolverlo.

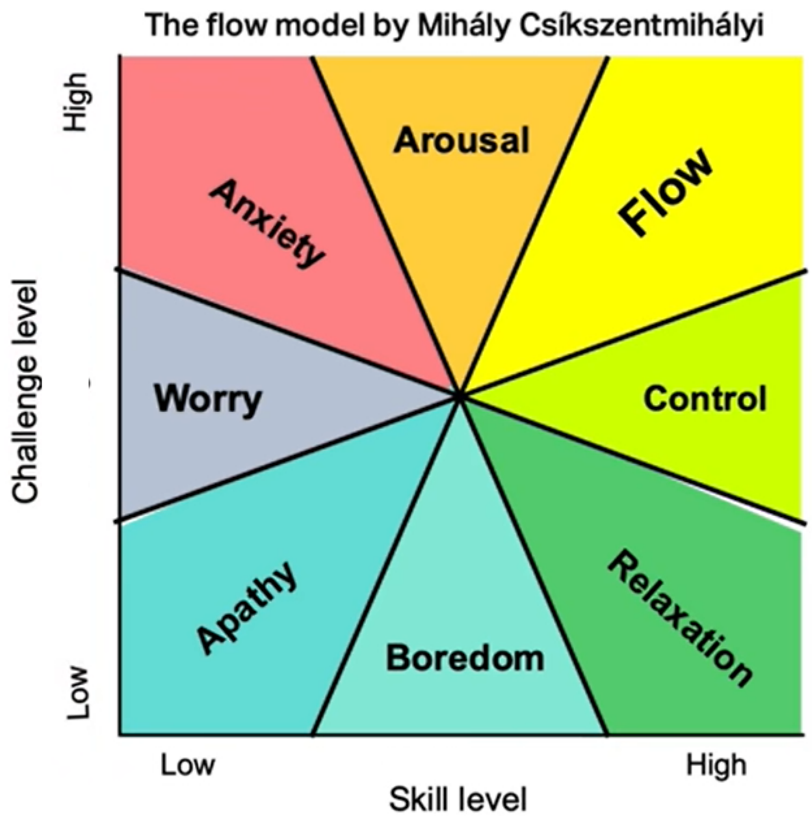


Figura E.9: *Skill vs challenge level*

65. El 50 % del tiempo se destina a salvar variables, comentar códigos, definir nombres correctos, hacer buenas estructuras en el código.

66. Procurar aprender algo nuevo cada día (videos de 5-10min al día), como:

    a) Ver videos de cómo hacer gráficas en R

    b) % % > % % en R para filtrar información en matrices

    c) Excel

    d) Cosas de R

67. Tener en cuenta que el código hace cosas similares por materia. Se tienen los mismos errores por materia.

68. Un programa de computadora, con que haga los mismos errores que un humano, es bueno porque su costo es menor.

69. Se espera que después de la primera generación, la mejora en las calificaciones de las asignaciones sea grande. Casi es una regla para el AG.

70. No poner cosas de otros si no se les menciona. Se se pone => hacer referencia/darle crédito.

71. No se da 2 veces la misma materia al mismo profesor para que los alumnos tengan mayor gama de profesores para elegir.

72. Mientras más información se pueda imprimir para darse cuenta de los posibles errores, es muy útil.
73. X_4 : Analizar presentación: Hacer varias pruebas con distintas combinaciones y elegir el mejor estilo/presentación.
74. X_{14} : Revisar/Investigar al respecto del problema y resolverlo.
75. 29/12/20202: Ya tenemos la herramienta, ahora hay que hacer pruebas para entenderla y explicarle a los demás su funcionamiento.
76. 29/12/20202: Actualmente falta por asignar $\frac{1}{3}$ de los grupos. La meta es que sólo falte el 10 % de los grupos (o menos).
77.
78.
79.
80.
81.
82.
83.
84.
85.
86.
87.
88.
89.
90.

Bibliografía

[1] Breusch T. S. y Pagan R., (1979), *A Simple Test for Heteroscedasticity and Random Coefficient Variation*, Ecometrica, Vol. 47, No. 5, pp. 1287 - 1294

[2] Casella G., (2006), *Statistical Inference*, Thomson Press

[3] Chatfield C. y Xing H., (2019), *The Analysis of Time Series An Introduction with R*, Chapman & Hall/CRC

[4] Do Chuong B. y Batzoglou S., (2008), *What is the expectation maximization algorithm?*, Nature Biotechnology, Vol. 26, No. 8, pp. 897 - 899

[5] Cox D. R. y Stuart A., (1955), *Some Quick Sign Tests for Trend in Location and Dispersion*, Biometrika, Vol. 42, No. 1/2, pp. 80 - 95

[6] Gibbons J. D. y Chakraborti S., (2011), *Nonparametric Statistical Inference*, Chapman & Hall/CRC

[7] Jarque C. M. y Bera A. K., (1980), *Efficient tests for normality, homoscedasticity and serial independence of regression residuals*, Economic Letters, Vol. 6, No. 3, pp. 255 - 259

[8] Lytras D., (2015), *On Seasonality: Comparing X-13ARIMA-SEATS Diagnostics for Quarterly Series*, U.S. Census Bureau

[9] Madsen H., (2008), *Time Series Analysis*, Chapman & Hall/CRC

[10] Miller L. H., (1956), *Table of Percentage Points of Kolmogorov Statistics*, Journal of the American Statistical Association, Vol. 51, No. 273, pp. 111-121.

[11] Montgomery D., Jennings C. y Kulahci M., (2015), *Introduction to Time Series Analysis and Forecasting*, Wiley

[12] Reeves C. R. y Rowe J.E., (2007), *Genetic Algorithms: Principles and Perspectives. A Guide to GA Theory*, Kluwer Academic Publishers

[13] Rincón L., (2007), *Curso intermedio de probabilidad*, UNAM

[14] Rubinstein R. y Kroese D., (2016), *Simulation and the Monte Carlo Method*, Wiley

[15] Shumway R. y Stoffer D., (2017), *Time Series Analysis and Its Applications: With R Examples*, Springer

[16] Sivanandam S. N, y Deepa S. N,, (2008), *Introduction to Genetic Algorithms*, Springer

[17] Vazquez J., Naranjo L., Fuentes R. y Chávez M., (2018), *Introducción a la Estadística*, Proyecto PAPIME UNAM PE107117

[18] Yazdani M., Naeri B. y Zeinali E., (2017), *Algorithms for university course scheduling problems*, Tehnički vjesnik, Vol. 24, No. 2, pp. 241-247