



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

INFERENCIA ESTADÍSTICA APLICADA PARA
GENERAR UNA PROPUESTA DE GRUPOS PARA
LAS CARRERAS DE ACTUARÍA, CIENCIAS DE
LA COMPUTACIÓN, MATEMÁTICAS Y
MATEMÁTICAS APLICADAS DE LA FACULTAD
DE CIENCIAS

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIA

PRESENTA:
MIRIAM GABRIELA COLÍN NÚÑEZ

TUTOR
DR. ARRIGO COEN CORIA

2021





UNAM - Dirección General de Bibliotecas

Tesis Digitales

Restricciones de uso

DERECHOS RESERVADOS ©

PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

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:

Mat.
Margarita Elvira
Chávez
Cano

Datos del sinodal 2:

Dra.
Lizbeth
Naranjo
Albarrán

Datos del sinodal 3:

Dra.
Ruth Selene
Fuentes
García

Datos del sinodal 4:

M. en C.
Jaime
Vázquez
Alamilla

Datos del sinodal 5:

M. en C.
Gustavo Adolfo
García
Cano

Datos del sinodal 6:

Dra.
Ana
Meda
Guardiola

Datos del trabajo escrito:

Inferencia estadística aplicada para generar una propuesta de grupos para las carreras de Actuaría, Ciencias de la Computación, Matemáticas y Matemáticas Aplicadas de la Facultad de Ciencias

(Número de Páginas) p
2021

Dedicado a

Agradecimientos

¡Muchas gracias a todos!

Índice general

1. Introducción	1
1.1. Motivación	1
1.2. Definición de conceptos	2
1.3. Nomenclatura	2
1.4. Planteamiento del problema	2
1.5. Objetivos	4
1.6. Datos a analizar	4
1.6.1. Análisis por tipo de semestre: par e impar	5
1.6.2. Análisis por turno: matutino y vespertino	6
2. Extracción de datos	9
2.1. Estructura de las URL	10
2.2. Extracción de datos con la aplicación SelectorGadget	11
2.3. Tipos de grupos en las páginas web de la Facultad de Ciencias	12
2.4. Limpieza de base de datos	14
2.4.1. Problemas de falta de información	14
2.4.2. Problemas de información repetida	16
2.4.3. Otros problemas al extraer información	19
2.5. Matrices de datos	21
3. Análisis estadístico	25
3.1. Análisis estadístico básico	26
3.1.1. Prueba de tendencia	28
3.1.2. Prueba de estacionalidad	29
3.1.3. Prueba de homocedasticidad	30
3.2. Análisis estadístico por grupo de datos	32
3.3. Análisis estadístico por carrera	34
3.4. Distribución del tamaño de los grupos	36
3.5. Comportamientos por hora	40
4. Simulación	43
4.1. Obtención de nombres de materias	44
4.2. Obtención de los parámetros q_1 y q_2	45
4.3. Obtención de nombres de profesores	49
4.3.1. Profesores de tiempo completo	50
4.3.2. Profesores de asignatura	52

4.4. Simulación de tamaño de grupos	53
4.5. Simulación de solicitudes de profesores	53
4.6. Simulación de la demanda de alumnos	55
4.7. Modelo de Mezcla Gaussiana	57
4.8. Obtención de D' y D_0	60
4.9. Simulación de esqueletos	64
4.10. Obtención de mat_esqueleto	64
5. Algoritmo Genético	69
5.1. Algoritmo Genético aplicado a los horarios	70
5.2. Resultados del Algoritmo Genético	73
6. Conclusiones	81
Apéndice A. Materias agrupadas	83
Apéndice B. Resultados útiles	87
Apéndice C. Abreviaturas	89
Apéndice D. Ejemplo de asignación final	91
Apéndice E. Observaciones / Notas	115
Bibliografía	127

Índice de figuras

1.1.	<i>Número de alumnos por semestres pares e impares: Probabilidad I</i>	5
1.2.	<i>Histogramas del número de alumnos por semestre: Probabilidad I</i>	6
1.3.	<i>Número de alumnos por turno: Probabilidad I</i>	7
1.4.	<i>Histogramas del número de alumnos por turno: Probabilidad I</i>	7
2.1.	<i>Página de horarios de la Facultad</i>	9
2.2.	<i>Uso de la aplicación SelectorGadget</i>	12
2.3.	<i>Tipo de grupo A</i>	13
2.4.	<i>Tipo de grupo B</i>	13
2.5.	<i>Tipo de grupo C</i>	13
2.6.	<i>Página web en blanco</i>	14
2.7.	<i>Grupo sin información de salón</i>	15
2.8.	<i>Grupo sin información de alumnos</i>	15
2.9.	<i>Grupo sólo con horario</i>	16
2.10.	<i>Información repetida: Planes de estudio</i>	17
2.11.	<i>Información repetida: Materia con nombres distintos</i>	18
2.12.	<i>Información repetida: Mismo profesor, misma hora, materias distintas</i>	19
2.13.	<i>Grupo con un alumno</i>	20
2.14.	<i>Grupo con medias horas</i>	20
2.15.	<i>Grupo con horarios múltiples</i>	20
2.16.	<i>Grupo de inglés</i>	21
2.17.	<i>Grupo con estructura diferente</i>	21
3.1.	<i>Descomposición por el método de promedios móviles</i>	27
3.2.	<i>Media de alumnos por semestre</i>	28
3.3.	<i>Prueba Cox-Stuart para aleatoriedad</i>	29
3.4.	<i>Prueba Cox-Stuart para tendencia</i>	29
3.5.	<i>Número total de alumnos por semestre</i>	30
3.6.	<i>Prueba QS para estacionalidad</i>	30
3.7.	<i>Desviación estándar del número de alumnos por semestre</i>	31
3.8.	<i>Prueba Jarque-Bera para normalidad</i>	31
3.9.	<i>Prueba Breusch-Pagan para homocedasticidad</i>	32
3.10.	<i>Número de alumnos de semestres pares e impares</i>	33
3.11.	<i>Histogramas del número de alumnos en semestres pares e impares</i>	33
3.12.	<i>Número de alumnos por turno de todos los semestres</i>	34
3.13.	<i>Histogramas del número de alumnos en los turnos matutino y vespertino</i>	35
3.14.	<i>Histogramas del número de alumnos por carrera</i>	35

3.15. Densidades del número de alumnos por carrera	36
3.16. Histograma del número de alumnos por grupo de todos los semestres	37
3.17. Densidades del número de alumnos por grupo de cada semestre	38
3.18. Histograma con densidades ajustadas	39
3.19. Número promedio de grupos por hora	41
3.20. Número promedio de alumnos por hora	41
4.1. Diagrama de flujo de la función “gen_asignacion()”	44
4.2. Matriz con información por materia	46
4.3. Promedio de la desviación estándar: 5 materias, 12 intervalos	47
4.4. Promedio de la desviación estándar: 10 materias, 6 intervalos	47
4.5. Promedio de la desviación estándar: 10 materias, 4 intervalos	48
4.6. Promedio de la desviación estándar: 5 materias, 4 intervalos	48
4.7. Diagrama de los intervalos de confianza	48
4.8. Matriz con medidas de dispersión de prueba aleatoria	49
4.9. Profesores de tiempo completo: SelectorGadget	50
4.10. Vector de profesores de tiempo completo	51
4.11. Ejemplo de matriz de solicitudes de un profesor	54
4.12. Ejemplo de matriz con alumnos corregidos	55
4.13. Ejemplo de vector con demanda simulada para el 2020-2 de “Modelos de Supervivencia y Series de Tiempo”	56
4.14. Ejemplo de matriz con demanda simulada para el 2020-2	57
4.15. Histograma del número de alumnos esperados por hora: Modelo inicial de mezcla de normales	59
4.16. Histograma del número de alumnos esperados por hora: Modelo final de mezcla de normales	59
4.17. Metodología A	61
4.18. Metodología B	61
4.19. Metodología C	62
4.20. Metodología D	62
4.21. Heatmap metodología B	63
4.22. Heatmap metodología C	63
4.23. Histograma con los datos del esqueleto inicial	65
4.24. Histograma con todos los datos de los esqueletos simulados	66
4.25. Ejemplo de esqueleto para el semestre 2020-2	67
5.1. Diagrama del Algoritmo Genético	70
5.2. Calificaciones de mejores asignaciones	74
5.3. Gráficas de caja de calificaciones de asignaciones por generación	75
5.4. Media de calificaciones de asignaciones por generación	76
5.5. Varianza de calificaciones de asignaciones por generación	76
5.6. Gráficas de caja del número de genes en asignaciones por generación	77
5.7. Solicitudes de la materia “Teoría del Seguro”	79
E.1. Resumen de clases de inglés antes de modificación	116
E.2. Ejemplo de horarios de semestre 2021-1	118
E.3. Notas de T26	119
E.4. Ejemplo de Roxygen	119

E.5. <i>Ejemplo de varianza</i>	120
E.6. <i>Cláusula 99 CCTPA: Ayuda para la impresión de la tesis</i>	121
E.7. <i>Nombres planes de estudio</i>	122
E.8. <i>which in plot</i>	123
E.9. <i>Skill vs challenge level</i>	124

Índice de tablas

1.1. <i>Ejemplo de asignación</i>	4
1.2. <i>Grupos de datos</i>	8
2.1. <i>Planes de estudio por carrera con clave</i>	10
2.2. <i>Descripción de las columnas de la matriz “mat_posibles_url”</i>	11
2.3. <i>Descripción de las columnas de la matriz “m_grande”</i>	23
4.1. <i>Posibles valores para q_1 y q_2</i>	45
4.2. <i>Diferencias en nombres de profesores de tiempo completo</i>	51
4.3. <i>Diferencias en nombres de profesores de asignatura</i>	52
5.1. <i>Resumen de las pruebas del Algoritmo Genético</i>	74
5.2. <i>Submatriz con asignación final</i>	78
C.1. <i>Abreviaturas</i>	89
D.1. <i>Matriz con asignación final</i>	113

Códigos

E.1. <i>Ejemplo de ciclo for</i>	116
E.2. <i>Ejemplo de estructura de funciones</i>	117

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. Además, el modelo propuesto puede ser aplicado para resolver problemas generales de asignación de recursos.

Actualmente para hacer la asignación de horarios primero se reúne 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: Matriz con el número de grupos por hora y por materia.

Asignación: Matriz con las columnas Materia-Profesor-Horario.

Grupo: Vector con las entradas Materia-Profesor-Horario.

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, $m = 202$.

p : Número de profesores considerados, $p = 1223$.

t : Número de horas en las que se imparten las materias, $t = 15$.

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

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

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

$U_{j,i,h}$: 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.

1.4. Planteamiento del problema

En el problema de asignación de horarios se quiere asociar un profesor con una materia y un horario. Existen trabajos que han abordado 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 d , más la utilidad de que la materia j sea impartida en el día d . 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 la materia j sea impartida por el profesor i a la hora h .

El planteamiento del problema maximización es el siguiente:

- 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):

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

- III) Restricciones:

$$\sum_{i=1}^p \sum_{h=1}^t x_{j,i,h} = 1 \quad \forall j \quad (1.1)$$

$$\sum_{j=1}^m x_{j,i,h} \leq 1 \quad \forall i, h \quad (1.2)$$

$$\sum_{h=1}^t x_{j,i,h} \leq V_{j,i} \quad \forall j, i \quad (1.3)$$

$$x_{j,i,h}, V_{j,i} \in \{0, 1\} \quad \forall j, i, h \quad (1.4)$$

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) especificamos que las variables utilizadas son binarias.

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 (h, j) contiene el número de grupos simulados de la hora h 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.
- Función calificadora de asignaciones: Califica con respecto al número de grupos simulados.

El conjunto de soluciones se presenta por medio de una matriz de tres columnas. El número de renglones depende del número de grupos se hayan simulado. En el k -ésimo renglón se tiene la información de la k -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
Modelos no Paramétricos y de Regresión	Jaime Vázquez Alamilla	10
Estadística Bayesiana	Ruth Selene Fuentes García	11
Modelos de Supervivencia y de Series de Tiempo	Lizbeth Naranjo Albarrán	13

Tabla 1.1: *Se muestra un ejemplo de una asignación. La matriz 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. El segundo criterio 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.

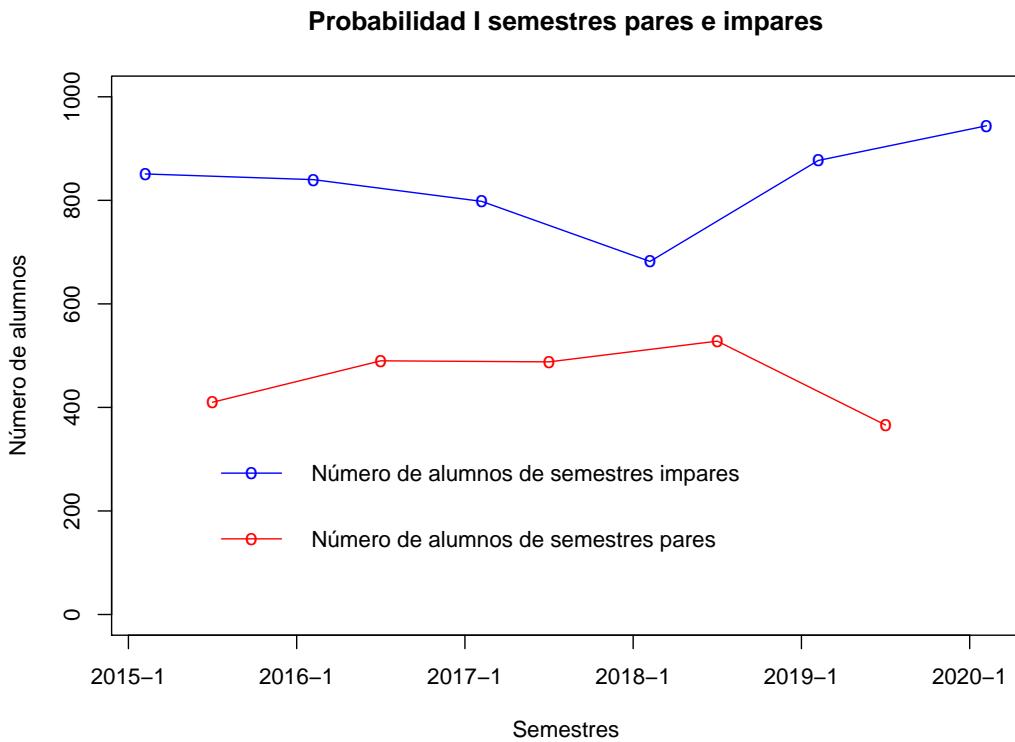


Figura 1.1: Se muestran las series de tiempo del número de alumnos por semestres (pares e impares) de “Probabilidad I”. Se puede observar que el número de alumnos de semestres impares es siempre mayor al de semestres pares.

La Figura 1.2 contiene dos histogramas, las barras rojas representan el número de alumnos por grupo de semestres pares. Las barras azules representan el número de alumnos por grupo de semestres impares.

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.

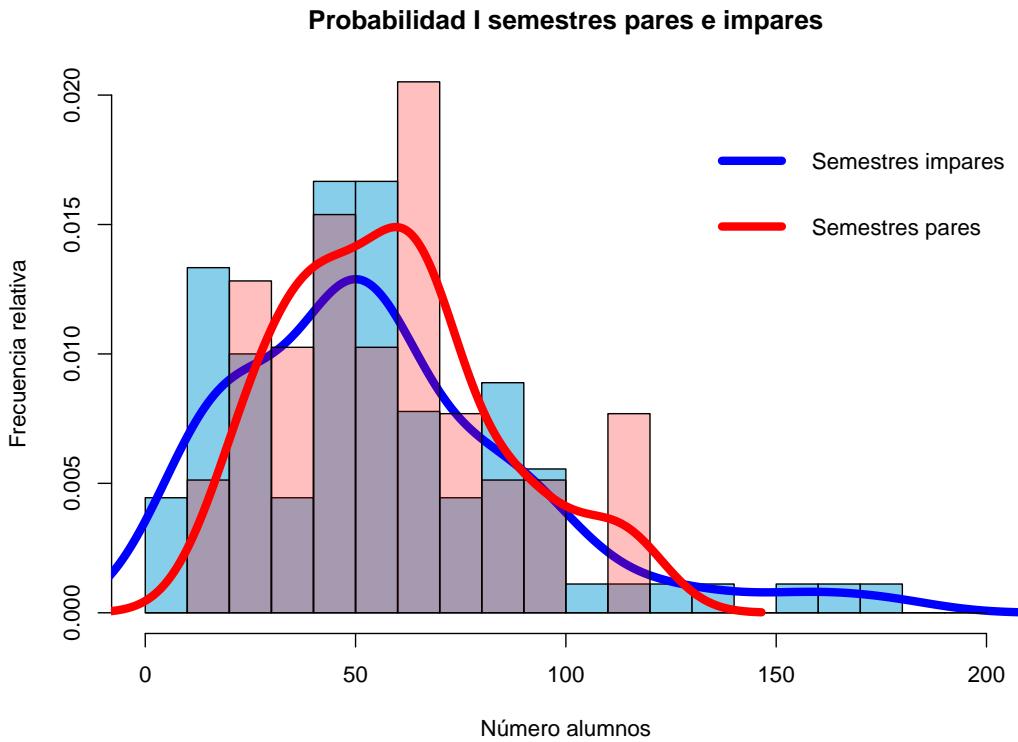


Figura 1.2: Se muestran los histogramas del número de alumnos por semestres (pares e impares) de “Probabilidad I”. Se puede observar que las densidades ajustadas son muy parecidas.

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.

También se puede ver que la varianza en el turno matutino es mucho mayor que en el turno vespertino. La desviación estándar del número de alumnos en el turno matutino es 179.20 y en el turno vespertino es 54.95. Ésto 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.

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.

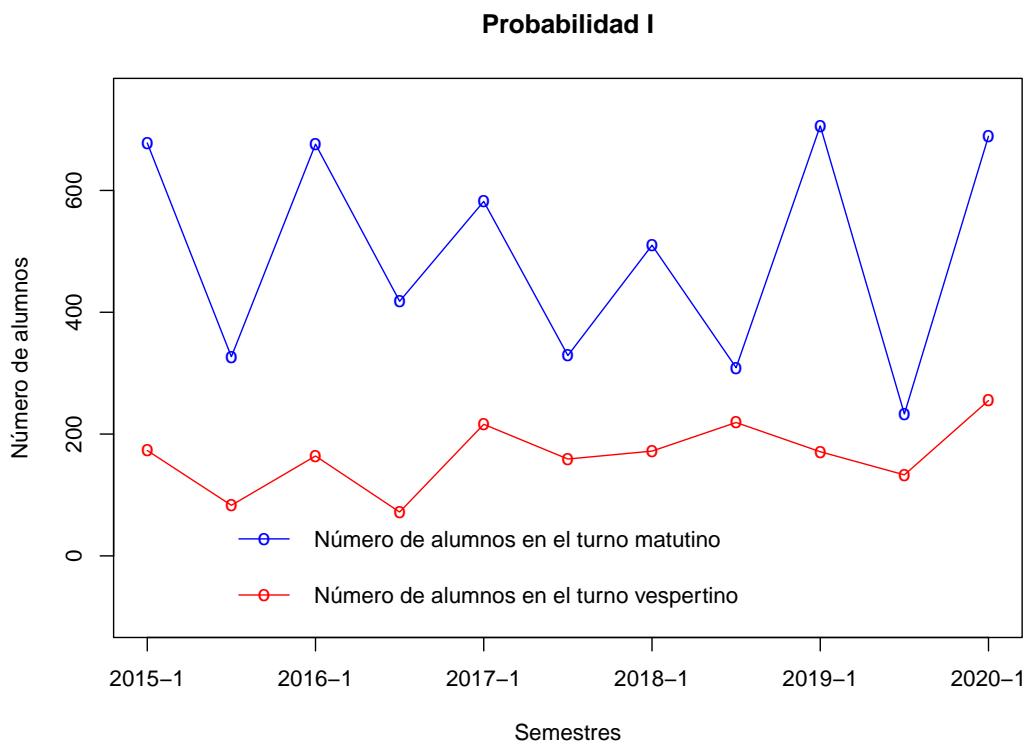


Figura 1.3: Se muestran las series de tiempo del número de alumnos por turno (matutino y vespertino) de “Probabilidad I”. Se puede ver que el número de alumnos del turno matutino es siempre mayor al número de alumnos del turno vespertino.

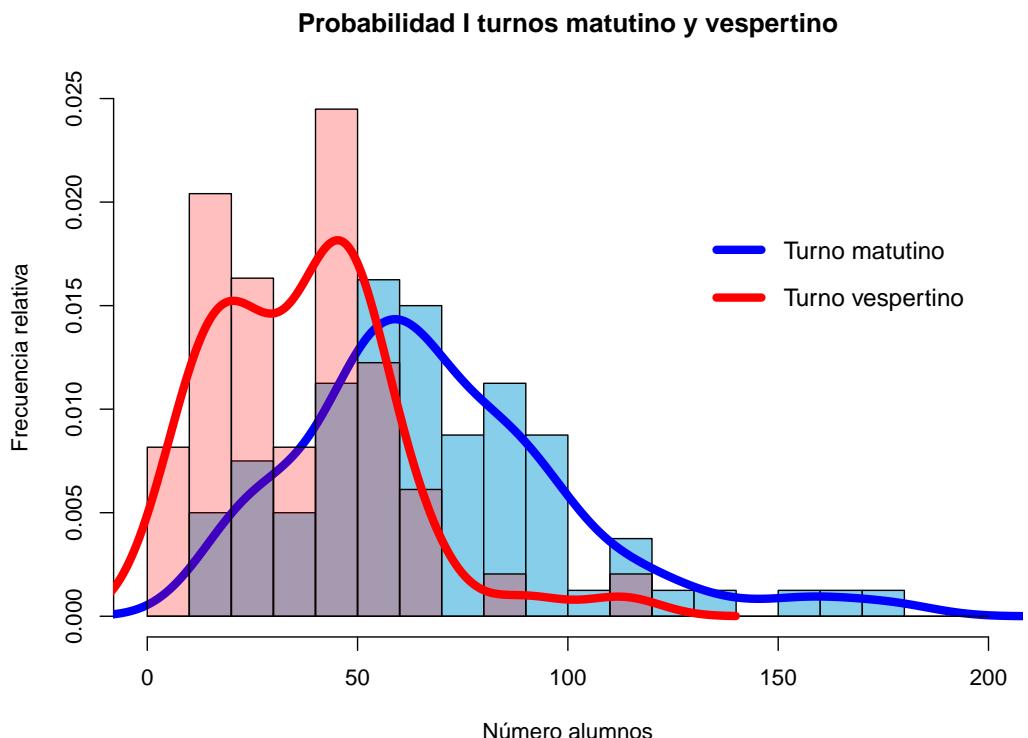


Figura 1.4: Se muestran los histogramas del número de alumnos por turno (matutino y vespertino) de “Probabilidad I”. Se puede observar que las densidades ajustadas son muy diferentes.

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).

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 una de ellas 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.

The screenshot shows the official website of the Faculty of Sciences at the National Autonomous University of Mexico (UNAM). The header features the UNAM logo and the text "Universidad Nacional Autónoma de México". Below the header, there is a navigation bar with links for "Inicio", "Contacto", "Mapa del sitio", "Directorio", "Correo", "Tienda Virtual", "Ingresar", and a search bar labeled "Buscar". The main content area is titled "Horarios 2020-1". On the left, there is a sidebar with links for "COMUNIDAD", "LICENCIATURA", "DOCENCIA", "INVESTIGACIÓN", "POSGRADO", "EXTENSIÓN", "SERVICIOS", "NOSOTROS", and "EVENTOS". The main content area displays information for the "Matemáticas (plan 1983)" course. It shows two groups: Group 7068 (40 places, 37 students) and Group 7070 (40 places, 31 students). Each group has a list of professors, teaching assistants, and their respective schedules. For example, Group 7068 includes Profesor Fabio Ezequiel Miranda Perea (lunes a viernes 11 a 12 O125), Ayudante Javier Enriquez Mendoza (mañanas 11 a 12 O125), Ayud. Lab. Pablo Gerardo González López (miércoles 14 a 16 Laboratorio de Ciencias de la Computación 2), and a presentation. Group 7070 includes Profesor Fernando Abigail Galicia Mendoza (lunes a viernes 11 a 12 P211), Ayudante Pedro Juan Salvador Sánchez Pérez (mañanas 11 a 12 P211), Ayud. Lab. María Ximena Lezama Hernández (miércoles 14 a 16 Laboratorio de Ciencias de la Computación 1), and a presentation. The footer of the page includes the Faculty of Sciences logo and a "Facultad de Ciencias" link.

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

Al iniciar la búsqueda de información notamos que las URL 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 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 de las páginas de la Facultad.

Una vez identificada la estructura de las URL pudimos realizar la búsqueda de información de manera automatizada. Originalmente decidimos que $c \in \{1, 2, 3, \dots, 10000\}$. Después hicimos una función que genera una matriz llamada *mat_posibles_url*. La función sólo guarda las URL 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. Redujimos su conjunto de posibles valores y definimos $c \in \{1, \dots, 1000\}$.

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	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	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 $25 \times c = 25,000$. Este valor es una aproximación del número de posibles URL 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.

Deseamos obtener información de cada una de esas páginas. Obtener dicha información ingresando a cada una de ellas 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 coinciden (en color amarillo).
4. En caso de que se haya seleccionado más información de la deseada, 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 seleccionadas (en óvalos rojos).

ciencias.unam.mx/docencia/horarios/20181/2017/625

Actuaría (plan 2015)

Probabilidad I, Tercer Semestre

Buscar en grupos

Grupo 9017 51 alumnos.

Exámenes finales lunes 8 de enero 2018 y lunes 15 de enero 2018 de 9 a 11 en el O219.

Profesor Juri Salazar Flores	lu mi vi 9 a 10	O219
Ayudante Elvia Jimena Vidrio Hernández	ma ju 9 a 10	O219
Ayudante Even Zuriel Torres Villegas	ma ju 9 a 10	

Grupo 9018 40 alumnos.

Exámenes finales miércoles 10 de enero 2018 y miércoles 17 de enero 2018 de 9 a 11 en el O218.

Profesor Fernando Baltazar Larios	lu mi vi 9 a 10	O218
Ayudante Gabriela Stephanía Revueltas Hernández	ma ju 9 a 10	O218
Ayudante Abraham Ramírez Hernández	ma ju 9 a 10	

Grupo 9019 45 alumnos.

Exámenes finales lunes 8 de enero 2018 y lunes 15 de enero 2018 de 9 a 11 en el O217.

Profesor Maria Clara Fittipaldi	lu mi vi 9 a 10	O217
Ayudante Bruno Daniel Gómez Labra	ma ju 9 a 10	O217
Ayudante Brenda Liliana Transito Pimentel	ma ju 9 a 10	

Grupo 9020 33 alumnos.

Exámenes finales lunes 8 de enero 2018 y lunes 15 de enero 2018 de 9 a 11 en el O216.

Profesor Manuel Domínguez de la Iglesia	lu mi vi 9 a 10	O216
Ayudante Claudia Ivonne	tr:nth-child(1) td:nth-child(2)	a
Presentación		

Clear (15) Toggle Position ? X

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. 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.

- En el tipo A se tienen las páginas correspondientes al semestre actual. Este tipo de grupo tiene la información del número de lugares disponibles por salón, pero no contiene

2.3. TIPOS DE GRUPOS EN LAS PÁGINAS WEB DE LA FACULTAD DE CIENCIAS13

la información de los exámenes finales, porque se considera que el semestre 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 Jose 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-I).

- 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. Para cada caso se agregó el link correspondiente a la página web del ejemplo.

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>.

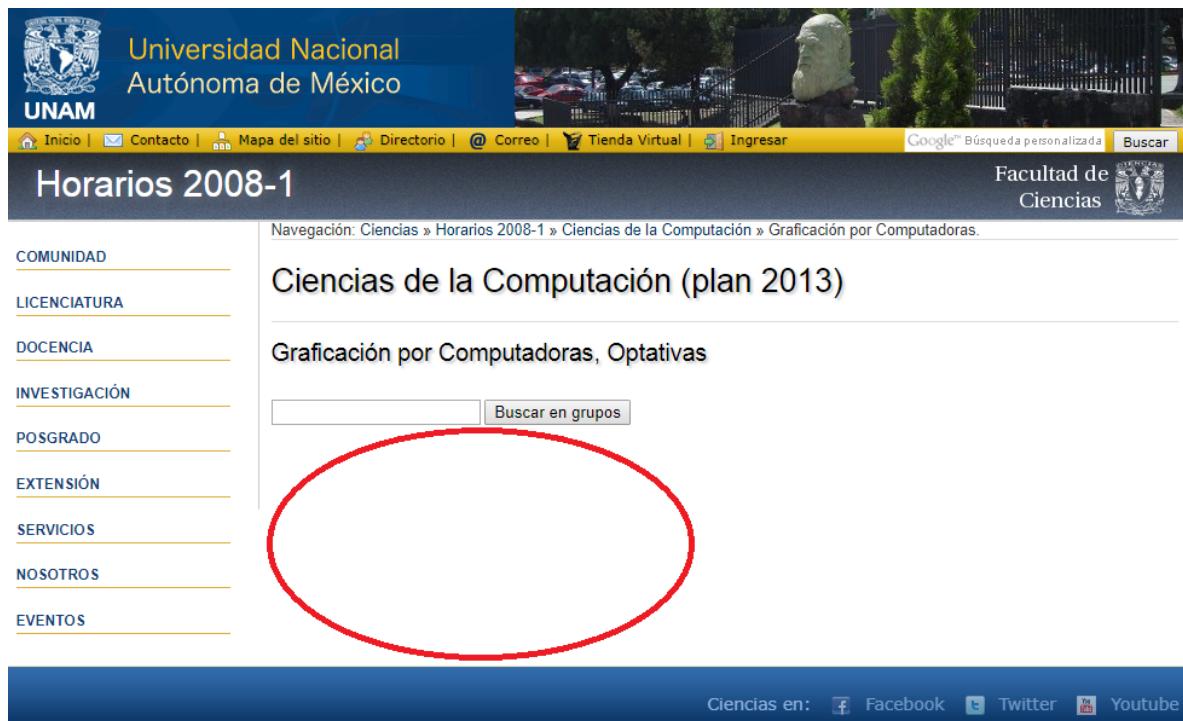


Figura 2.6: Se muestra un ejemplo de página web en blanco para la materia “Graficación por Computadoras”. En este tipo de páginas no encontramos información de ningún grupo.

- 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>.

Actuaría (plan 2000)

Álgebra Moderna IV, Optativas

Grupo 4250 6 alumnos.

Profesor José Ríos Montes lu mi vi 13 a 14

Ayudante ma ju 13 a 14



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>.

Actuaría (plan 2000)

Procesos Estocásticos I, Optativas

Grupo 6157 2 alumnos.

Profesor Fernando Guerrero Poblete lu mi vi 12 a 13 O216

Ayudante Héctor Alonso Olivares Aguayo ma ju 12 a 13 O216

Ayudante Rafael Martínez Sánchez ma ju 12 a 13

Ayudante Alfredo Hernández Lammoglia ma ju 12 a 13

Grupo 6192 3 alumnos.

Profesor Guillermo Garro Gómez lu mi vi 18 a 19 O122

Ayudante Martín Martínez Estrada ma ju 18 a 19 O122

Grupo 6193

Profesor Fernando Baltazar Larios lu mi vi 17 a 18 O221

Ayudante Estela Eréndira Zamora García ma ju 17 a 18 O221

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>.

Actuaría (plan 2000)

Variable Compleja II, Optativas

Grupo 4521 33 alumnos.

Profesor Guillermo Javier Francisco Sienra Loera lu mi vi 12 a 13 O123

Ayudante Adriana Andraca Gómez ma ju 12 a 13 O123

Presentación

Grupo 4519

Profesor lu mi vi 17 a 18

Ayudante ma ju 17 a 18

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>.

Horarios 2008-2

Ciencias de la Computación (plan 2013)

Graficación por Computadoras, Optativas

Buscar en grupos

Grupo 7054 19 alumnos.

Profesor Ana Luisa Solís González-Cosío lu mi vi 12 a 13

Ayudante José Israel Figueroa Angulo ma ju 12 a 13

Ayud. Lab. Azael Nieves Ramírez

(a) *Plan de estudios posterior*

Horarios 2008-2

Ciencias de la Computación (plan 1994)

Graficación por Computadoras, Optativas

Buscar en grupos

Grupo 7054 19 alumnos.

Profesor Ana Luisa Solís González-Cosío lu mi vi 12 a 13

Ayudante José Israel Figueroa Angulo ma ju 12 a 13

Ayud. Lab. Azael Nieves Ramírez

(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 Supervivencia 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. Cabe mencionar que ambas páginas corresponden al semestre 2020-1.

Matemáticas (plan 1983)

Estadística III, Optativas de los Niveles VII y VIII

Buscar en grupos

Grupo 9259, 35 lugares. 11 alumnos.

Profesor Claudia Lara Pérez Soto lu mi vi 9 a 10 101 (Nuevo Edificio)
Ayudante Ventura Jimenez Martinez ma ju 9 a 10 101 (Nuevo Edificio)

Grupo 9261, 81 lugares. 32 alumnos.

Profesor Sofia Villers Gómez lu mi vi 9 a 10 306 (Yelizcalli)
Ayudante Amílcar José Escobedo Pérez ma ju 9 a 10 306 (Yelizcalli)

Grupo 9263, 56 lugares. 9 alumnos.

Profesor Luis Antonio Rincón Solís lu mi vi 9 a 10 P102
Ayudante José Luis Miranda Olvera ma ju 9 a 10 P102

(a) *Matemáticas plan 1983*

Actuaría (plan 2015)

Modelos de Supervivencia y de Series de Tiempo, Séptimo Semestre

Buscar en grupos

Grupo 9258, 35 lugares. 11 alumnos.

Profesor Claudia Lara Pérez Soto lu mi vi 9 a 10 101 (Nuevo Edificio)
Ayudante Ventura Jimenez Martinez ma ju 9 a 10 101 (Nuevo Edificio)

Grupo 9260, 81 lugares. 32 alumnos.

Profesor Sofia Villers Gómez lu mi vi 9 a 10 306 (Yelizcalli)
Ayudante Amílcar José Escobedo Pérez ma ju 9 a 10 306 (Yelizcalli)

Grupo 9262, 56 lugares. 9 alumnos.

Profesor Luis Antonio Rincón Solís lu mi vi 9 a 10 P102
Ayudante José Luis Miranda Olvera ma ju 9 a 10 P102

(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*, del semestre 2011-1: <http://www.fciencias.unam.mx/docencia/horarios/20111/2017/162>.

En la subfigura (b) de la figura mencionada se muestra la información de la mate-

ria *Cálculo Diferencial e Integral I*, del semestre 2011-1: <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.

Actuaría (plan 2015)

Ecuaciones Diferenciales I, Cuarto Semestre

Grupo 1772 15 alumnos.

Profesor Edgar René Hernández Martínez lu mi vi 18 a 19 C123
 Ayudante Norma Angélica Cruz Cervantes ma ju 18 a 19 C123

(a) *Ecuaciones Diferenciales I*

Actuaría (plan 2015)

Cálculo Diferencial e Integral I, Primer Semestre

Grupo 4039 54 alumnos.

Profesor Edgar René Hernández Martínez lu a vi 18 a 19 Taller Interdisciplinario de Física y Biomedicina I
 Ayudante Norma Angélica Cruz Cervantes lu mi vi 19 a 20 Taller Interdisciplinario de Física y Biomedicina I
 Ayudante Luis Felipe Rivera Flores

(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: <http://www.fciencias.unam.mx/docencia/horarios/20081/119/1809>.

Grupo 6125 Un alumno.

Profesor Reyna Pineda González lu mi vi 21 a 22 102

Ayudante Elmo Jesús Viloria López ma ju 21 a 22 102

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. En la Figura 2.14 mostramos un ejemplo: <http://www.fciencias.unam.mx/docencia/horarios/20172/1556/820>. Se considera que esa materia inicia a las 18hrs.

Grupo 7014, 41 lugares. 19 alumnos.

Profesor Luis Alberto Ramírez Bermudez ma ju 18:30 a 20 Taller de Control y Electrónica

Ayudante Valente Vázquez Velázquez lu mi 20 a 21 Taller de Control y Electrónica

Ayud. Lab. Valente Vázquez Velázquez ju 14 a 16 Taller de Control y Electrónica

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. 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: <http://www.fciencias.unam.mx/docencia/horarios/20181/2055/1323>. 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.

Matemáticas Aplicadas (plan 2017)

Modelado y Programación, Investigación de Operaciones

	Buscar en grupos
--	------------------

Grupo 7035, 52 lugares. 44 alumnos.

Exámenes finales martes 9 de enero 2018 y martes 16 de enero 2018 de 13 a 15 en el O215.

Profesor José de Jesús Galaviz Casas lu mi vi 13 a 14 O215

Ayudante José Ricardo Rodríguez Abreu ma ju 13 a 14 O215

Ayud. Lab. Norma Verónica Trinidad Hernández ma 11 a 13 304 (Yelizcalli)

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. 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: <http://www.fciencias.unam.mx/docencia/horarios/20202/2017/1135>.

Grupo 9296, 45 lugares. 20 alumnos.
 Profesor Lilian Moreno Roldán sá 7 a 9 Sesión virtual
 ma 4 a 16 P207

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, debido a ello el código CSS utilizado no sirve para extraer 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: <http://www.fciencias.unam.mx/docencia/horarios/20201/2017/872>.

Actuaría (plan 2015)

Seminario de Investigación de Operaciones, Optativas

Grupo 9305 11 alumnos.
Juegos Evolutivos

Exámenes finales miércoles 27 de noviembre 2019 y miércoles 4 de diciembre 2019 de 10 a 12.

Profesor Claudia Villegas Azcorra lu mi vi 10 a 11 Grupo paralelo. Se impartirá en el Taller de Demografía.
 Ayudante Diego Eugenio Vallejo Carpintero ma ju 10 a 11 Grupo paralelo. Se impartirá en el Taller de Demografía.

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”, ...
2	Profesor	Nombre de la persona que va a impartir alguna materia	“Arrigo Coen Coria”, ...

La tabla continúa en la siguiente página

Col.	Nombre	Explicación	Posibles valores
3	Horario	Hora en la que se imparte alguna materia	“7 a 8”, ..., “21 a 22”
4	horario_num	Valores de la columna <i>Horario</i> en variables tipo <i>numeric</i>	7, 8, 9, ..., 20, 21
5	Lugares	Espacios disponibles por salón	N
6	Alumnos	Número de estudiantes inscritos por grupo	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	1, 2, 3, 4, 5
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	Dirección de una página web de los horarios de la Facultad correspondiente al grupo	URL 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
22	MatAp2017	Columna binaria, indica si el grupo pertenece a la carrera de MatAp, plan 2017	0, 1

La tabla continúa en la siguiente página

Col.	Nombre	Explicación	Posibles valores
23	NomMat_Act2000	Indica el nombre de la materia correspondiente a la carrera de Actuaría plan 2000	Nombres de materias de la Facultad
24	NomMat_Act2006	Indica el nombre de la materia correspondiente a la carrera de Actuaría plan 2006	Nombres de materias de la Facultad
25	NomMat_Act2015	Indica el nombre de la materia correspondiente a la carrera de Actuaría plan 2015	Nombres de materias de la Facultad
26	NomMat_CdC1994	Indica el nombre de la materia correspondiente a la carrera de CdC plan 1994	Nombres de materias de la Facultad
27	NomMat_CdC2013	Indica el nombre de la materia correspondiente a la carrera de CdC plan 2013	Nombres de materias de la Facultad
28	NomMat_Mat1983	Indica el nombre de la materia correspondiente a la carrera de Matemáticas plan 1983	Nombres de materias de la Facultad
29	NomMat_MAp2017	Indica el nombre de la 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 asignado (ver Sección 4.1)	1, 2, ..., 202

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.

Con las matrices *m_grande* de los semestres del 2008-1 al 2020-1, definimos la matriz *m_grande_total*. La información de esta matriz será utilizada a lo largo del trabajo.

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 dependencia de X_t con $\{X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_2, X_1, X_0, \dots\}$.

Denotamos a una serie de tiempo como:

$$X_t = m_t + s_t + y_t, \quad (3.1)$$

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:

$$X_t = m_t + s_t + y_t. \quad (3.2)$$

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:

$$X_t = m_t s_t y_t. \quad (3.3)$$

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:

$$X_t = m_t s_t + y_t. \quad (3.4)$$

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 períodos 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.

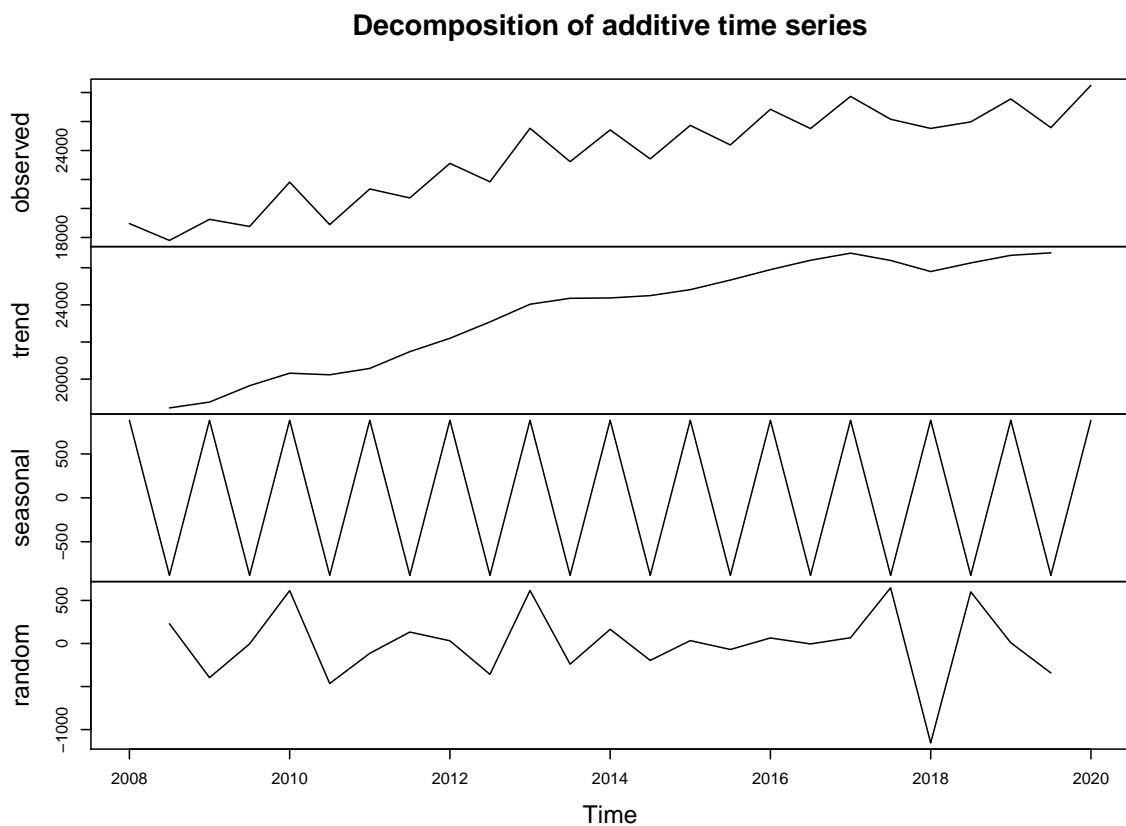


Figura 3.1: Se muestra la descomposición por el método de promedios móviles con la función “`decompose()`”. Los datos corresponden al número total de alumnos por semestre.

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.

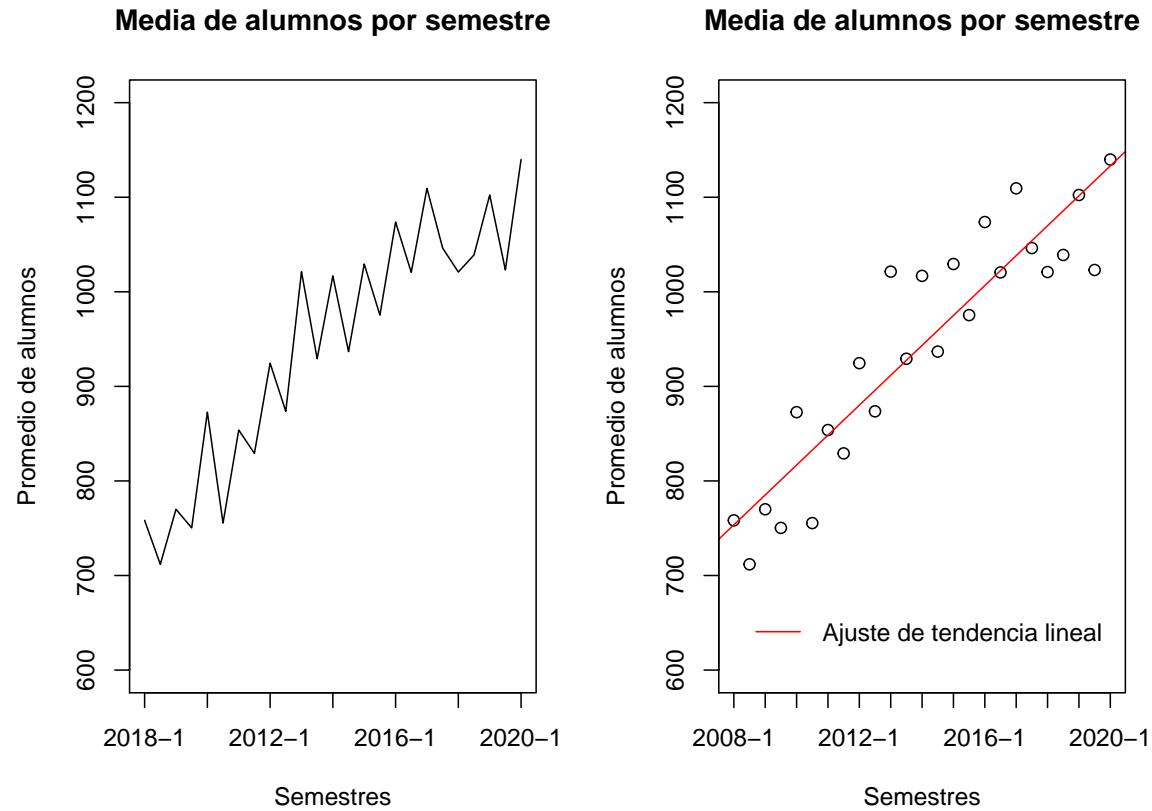


Figura 3.2: Se observa una tendencia creciente en la media de alumnos por semestre. La información corresponde a los semestres del 2008-1 al 2020-1.

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 $p\text{-value} \leq \alpha$, siendo α el nivel de significancia. Sea $\alpha = 0.01$. Como vemos en la Figura 3.3, $p\text{-value} = 0.0004 \leq 0.01 = \alpha$, 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 $p\text{-value} = 1 > 0.01 = \alpha$ 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*.

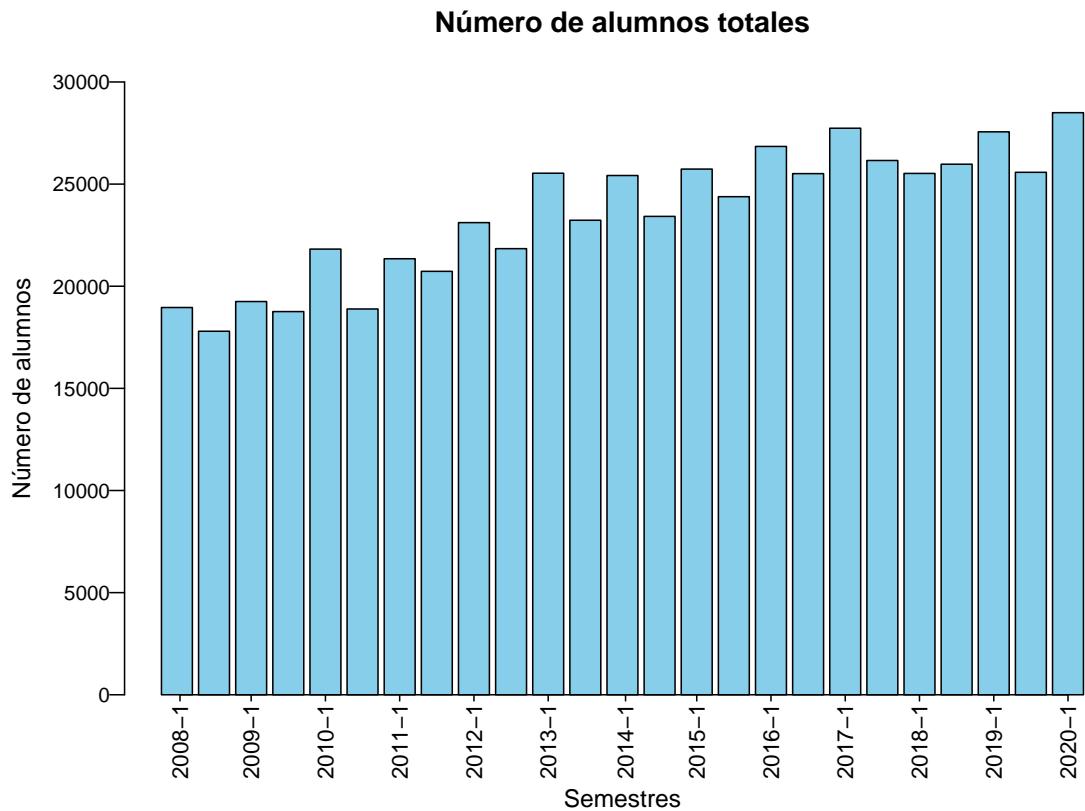


Figura 3.5: En esta figura se muestra la gráfica de barras del número total de alumnos inscritos por cada semestre. Se observa que año con año el número aumenta. En general, el número de alumnos de los semestres impares es mayor que el de su respectivo semestre par.

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 $p\text{-value} = 1.47 \times 10^{-6} \leqslant 0.01 = \alpha$ 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.

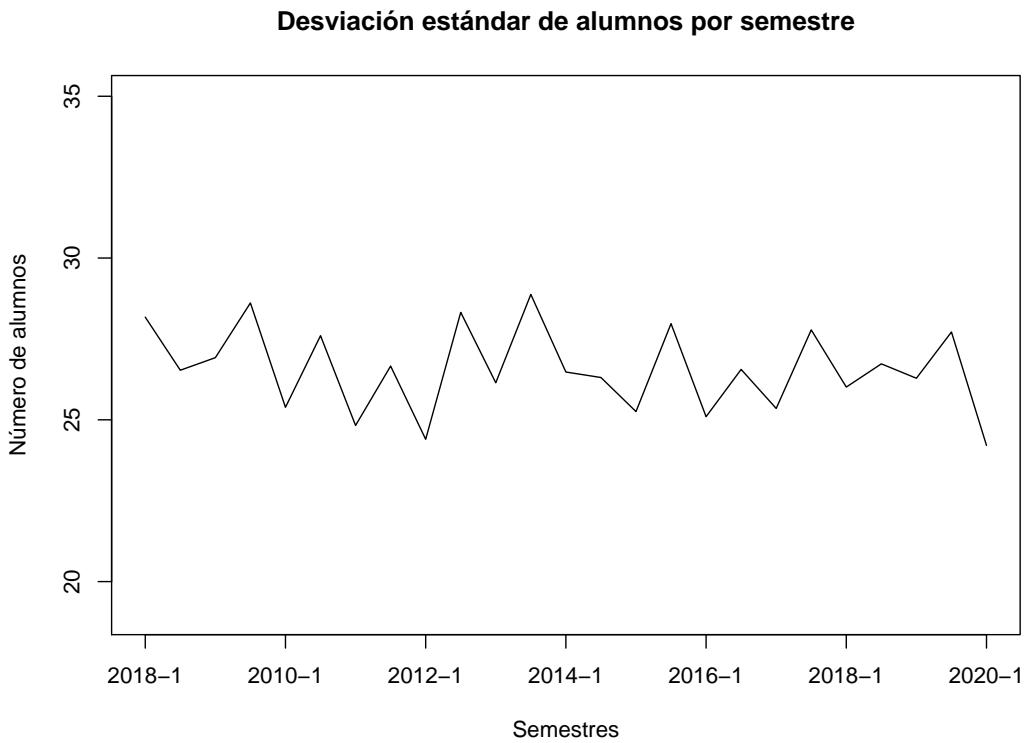


Figura 3.7: Se muestra la desviación estándar del número de alumnos por semestre. Se puede ver que el comportamiento de los datos es constante a lo largo del tiempo.

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 $p\text{-value} = 0.40 > 0.01 = \alpha$ 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 concluimos 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.

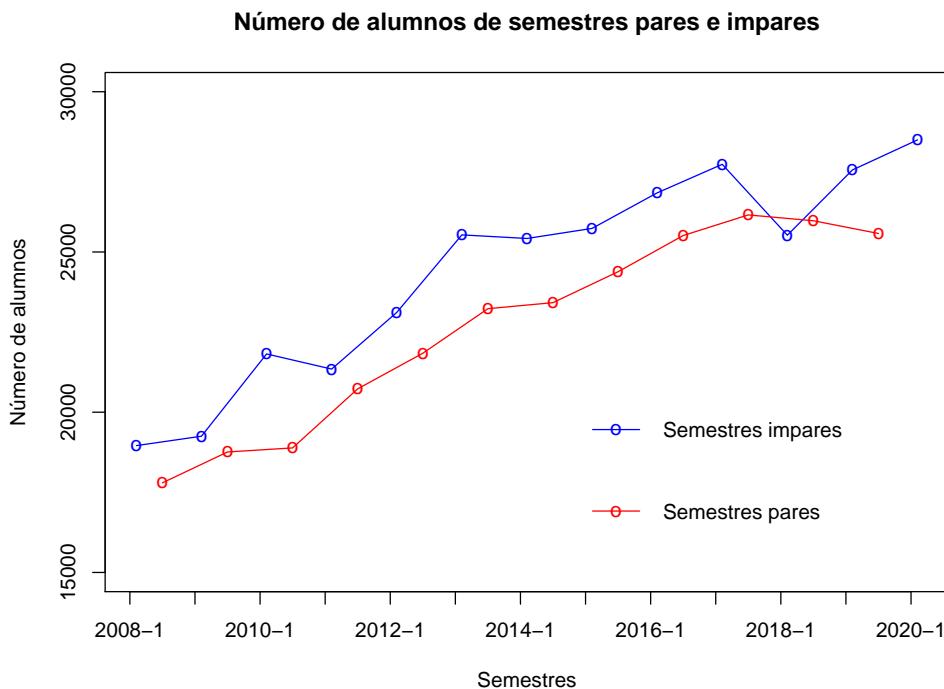


Figura 3.10: Se muestran las series de tiempo del número de alumnos en semestres pares e impares. Se observa una tendencia creciente y en general el número de alumnos de semestres impares (línea azul) es mayor al número de alumnos de semestres pares (línea roja).

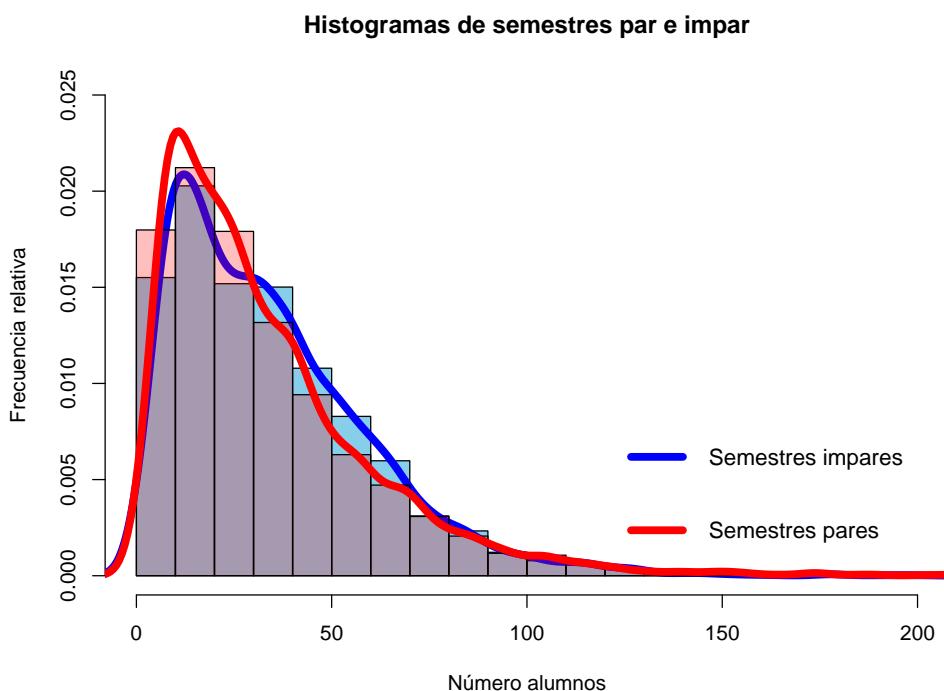


Figura 3.11: Se muestran los histogramas con el número de alumnos en semestres pares e impares. Las densidades ajustadas son muy parecidas, no importa si los datos son de semestres pares o impares.

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.

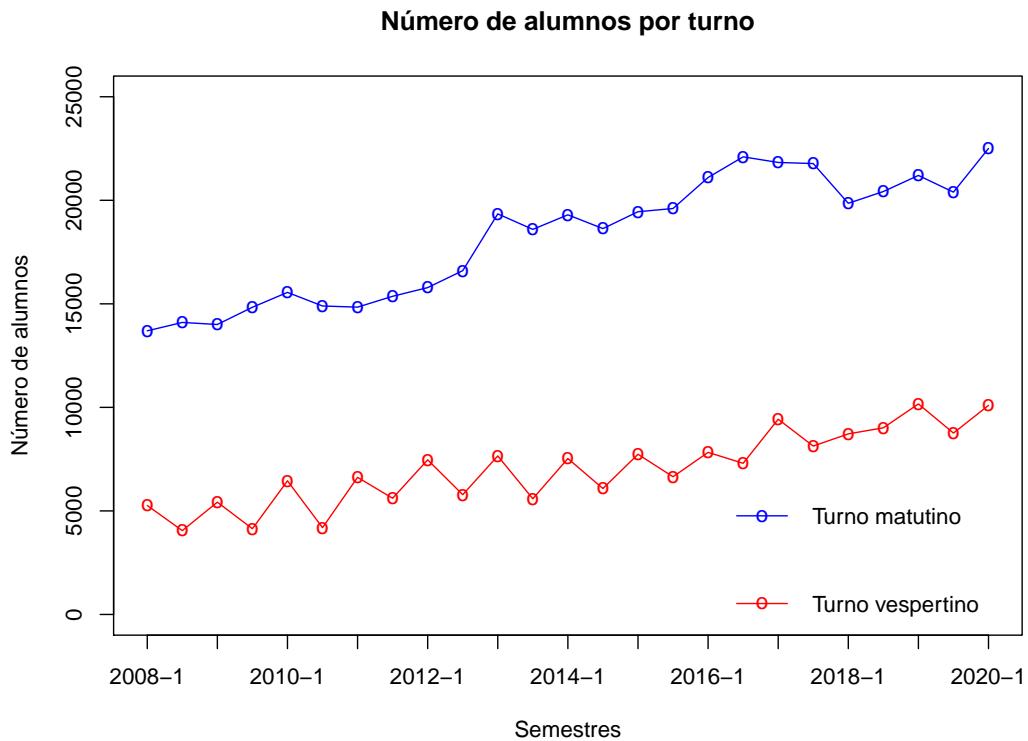


Figura 3.12: Se muestran las series de tiempo del número de alumnos por turno de todos los semestres. Se observa que el número de alumnos del turno matutino (línea azul) es siempre mayor al número de alumnos del turno vespertino (línea roja).

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

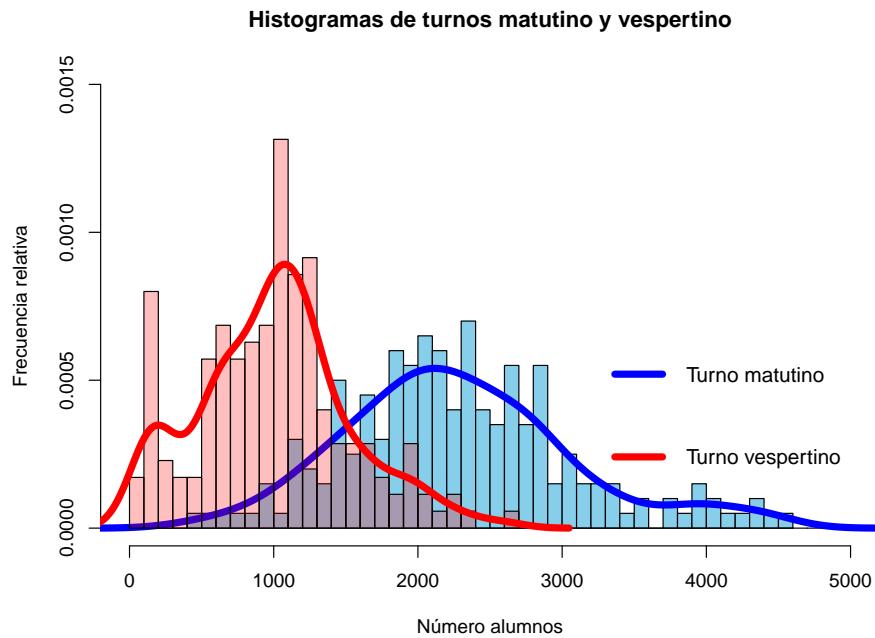


Figura 3.13: Se muestran los histogramas del número de alumnos en los turnos matutino y vespertino. Se puede concluir que hay más alumnos en el turno matutino que en el vespertino. Sus densidades ajustadas son diferentes.

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

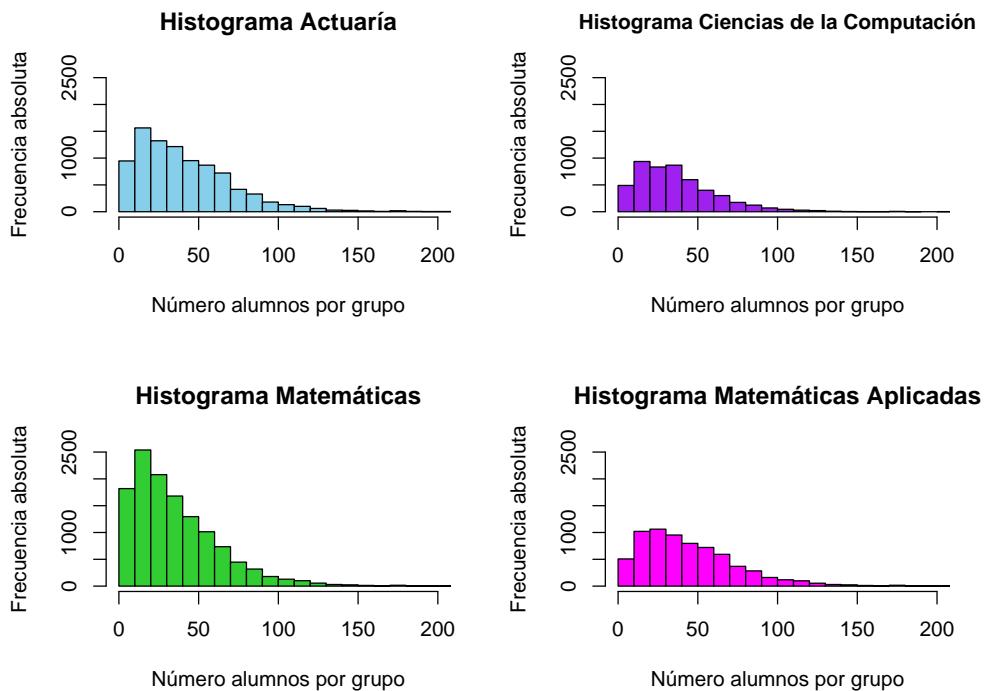


Figura 3.14: Se muestran los histogramas con el número de alumnos por grupo. Hay un histograma para cada carrera del Departamento de Matemáticas.

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.

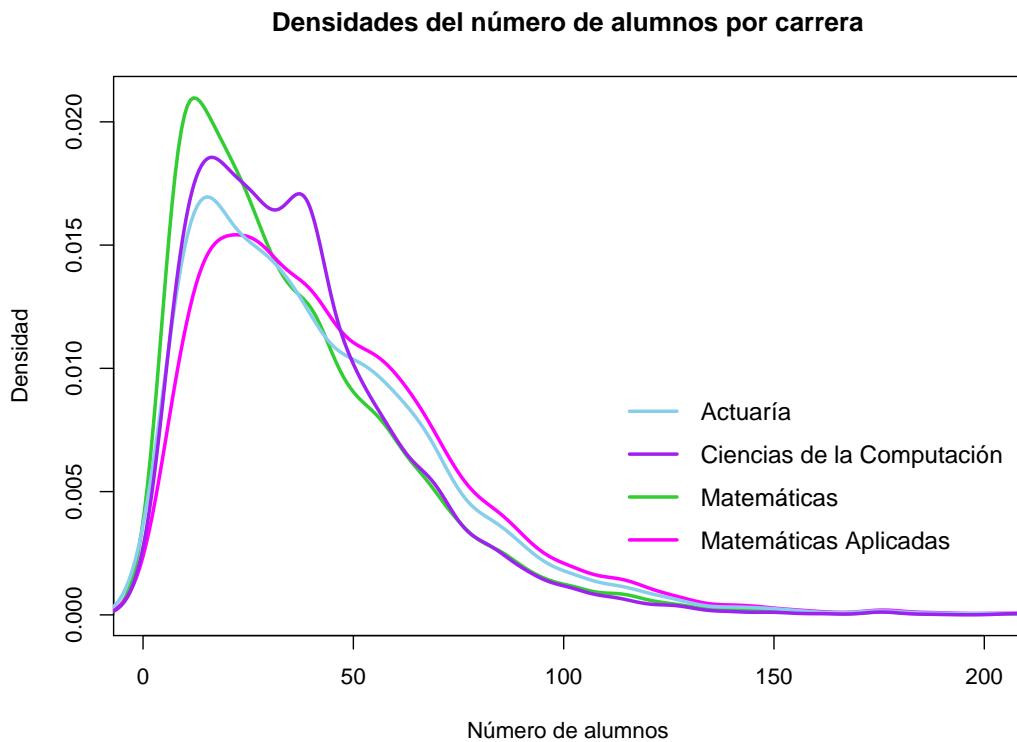


Figura 3.15: Se muestran las densidades ajustadas del número de alumnos por carrera del Departamento de Matemáticas.

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.

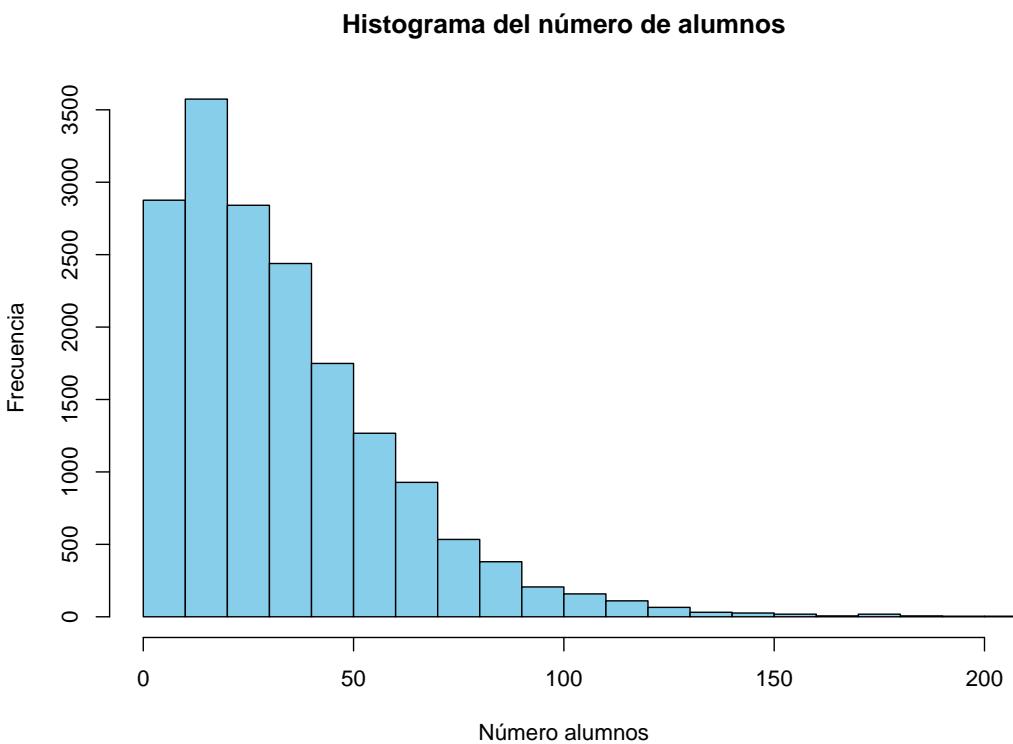


Figura 3.16: Se muestra el histograma del número de alumnos por grupo de todos los semestres. La información es de los semestres del 2008-1 al 2020-1. Vemos una mayor concentración en los grupos que tienen entre 10 y 40 alumnos.

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.

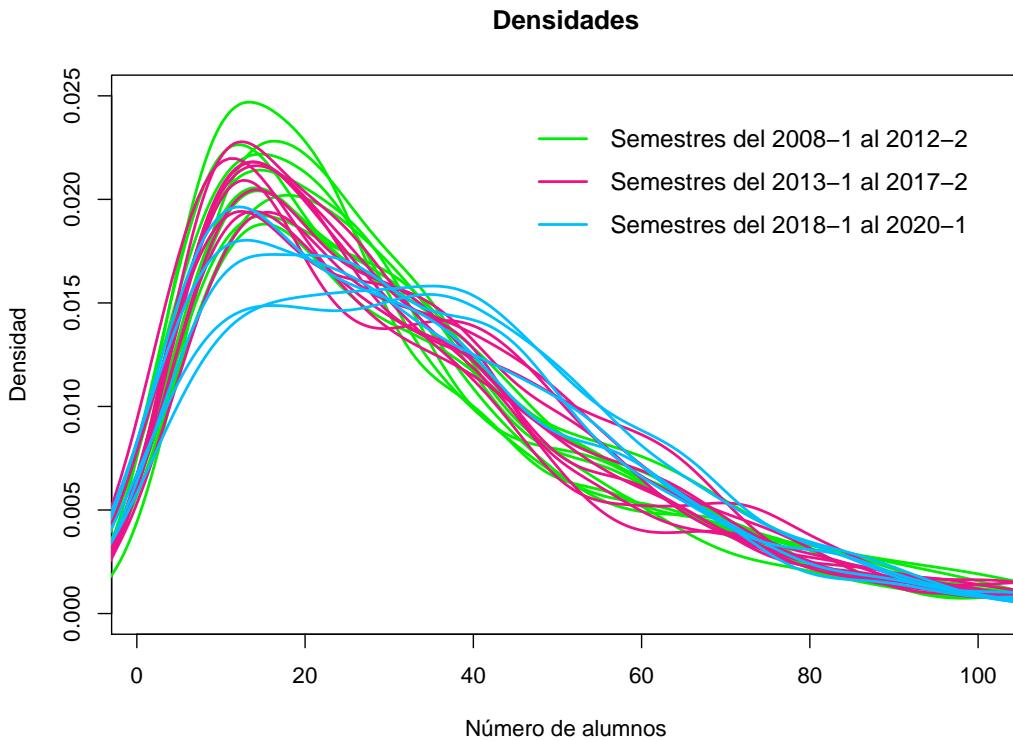


Figura 3.17: Se muestran las densidades ajustadas del número de alumnos por grupo de cada semestre. Cada línea corresponde a un semestre entre el 2008-1 y el 2020-1.

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:

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

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.

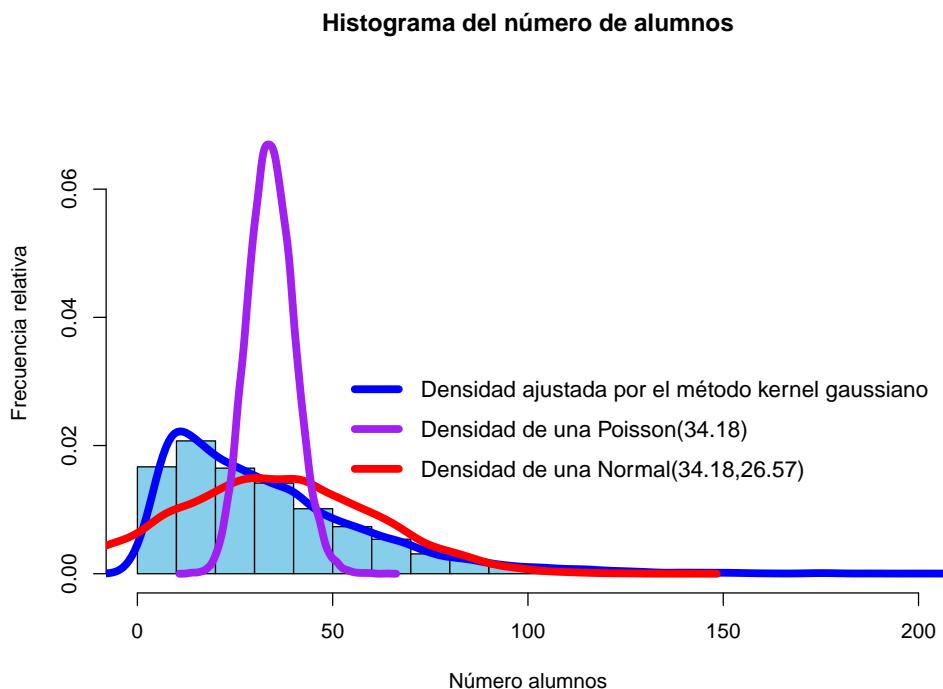


Figura 3.18: Se muestran 3 densidades ajustadas. La línea azul es la ajustada con la función “density()” en *R*. La morada es de una Poisson(34.18). La roja de una Normal(34.18,26.57). Ninguna de las distribuciones propuestas se ajustan de manera adecuada a los datos.

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 concluimos 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.

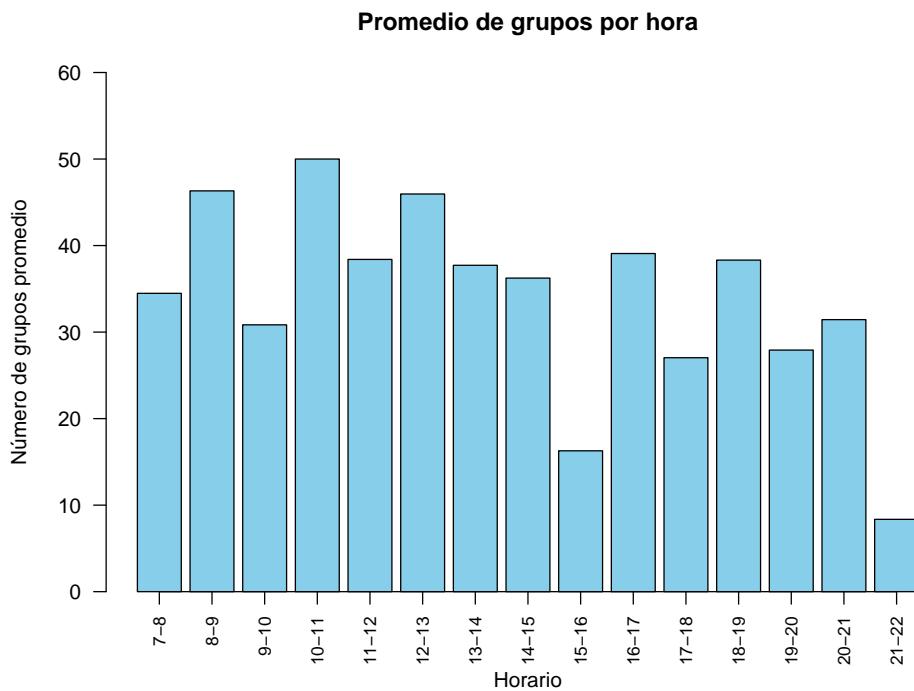


Figura 3.19: Se muestra la gráfica de barras del número promedio de grupos por hora. Se observa una disminución considerable a las 15hrs y a las 21hrs. El valor más alto se encuentra a las 10hrs.

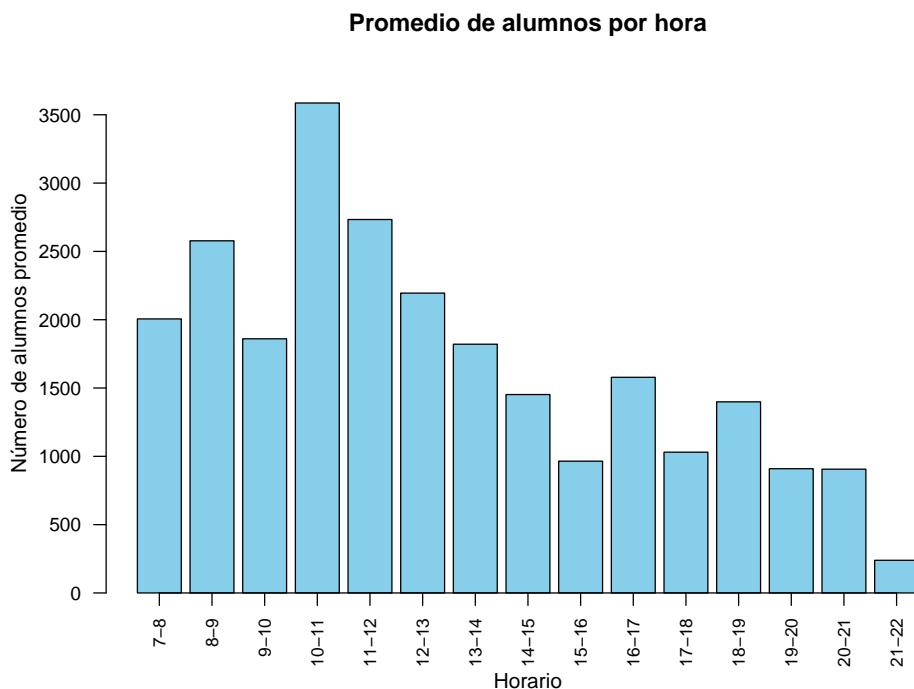


Figura 3.20: Se muestra la gráfica de barras del número promedio de alumnos por hora. Notamos una disminución de los valores a las 9hrs, a las 15hrs y a las 21hrs. El valor más alto lo encontramos a las 10hrs.

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. Esto 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.

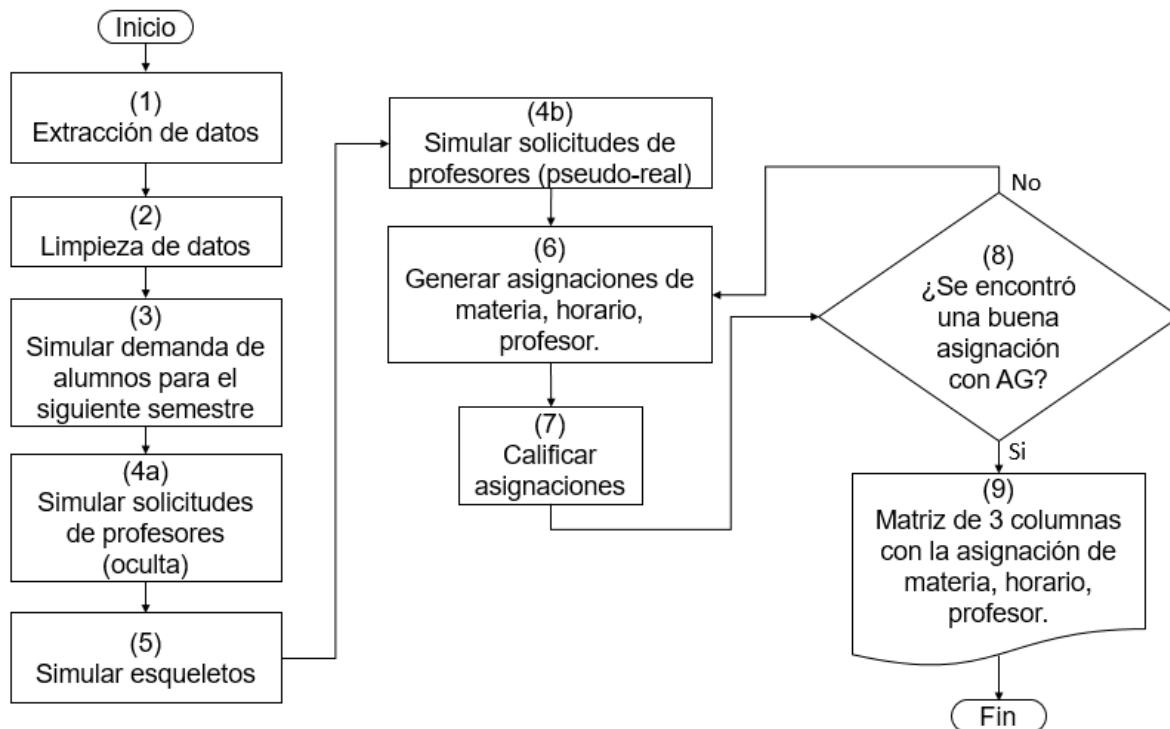


Figura 4.1: Se muestra el diagrama de flujo de la función “`gen_asignacion()`”. Se pueden ver los pasos para obtener la matriz de asignaciones.

4.1. Obtención de nombres de materias

Antes de iniciar las simulaciones definimos el vector `vec_nom_materias_total`. Éste contiene el nombre de las $m = 202$ materias encontradas en la matriz `m_grande_total` de los semestres 2008-1 al 2020-1. Dicho vector no tiene nombres repetidos. De esta manera evitamos 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 primera columna de dicha matriz, obtuvimos el vector `vec_nom_materias_total`. Los nombres del vector son los que utilizaremos para realizar las simulaciones de esqueletos y las asignaciones de horarios.

4.2. Obtención de los parámetros q_1 y q_2

Los valores de q_1 y q_2 representan los cuantiles utilizados al calcular los intervalos de confianza. Éstos son parámetros que se introducen en la función `hw()` de *R*. 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. Dicha función corresponde al método Holt-Winters aditivo.

Para seleccionar q_1 y q_2 , 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 estos 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 primera 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.

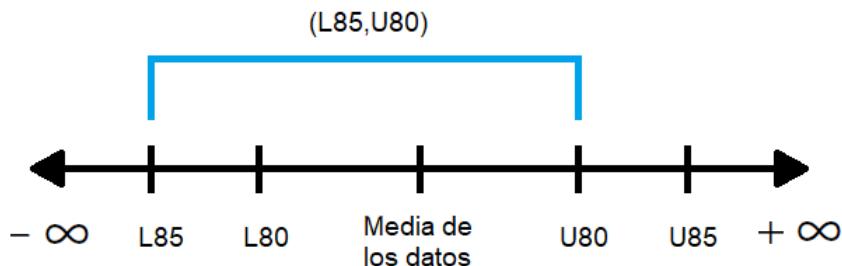


Figura 4.7: Se muestra un diagrama con el intervalo del que se va a obtener el número de alumnos para la simulación de cada materia en cada hora.

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

En la simulación de solicitudes se eligen materias y horarios con respecto a los profesores. Definimos la matriz *mat_nom_prof_total*, con 2 columnas. En ella guardamos información de cada profesor.

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.

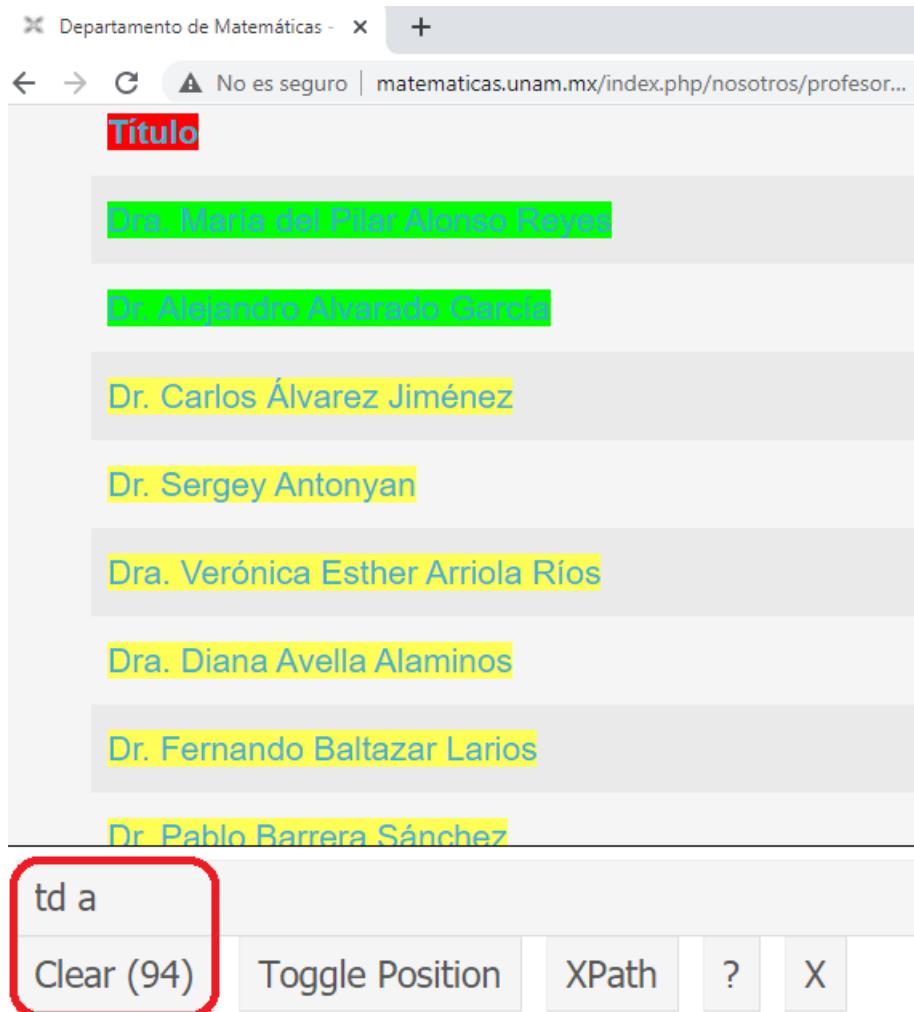


Figura 4.9: Se muestra la selección de profesores de tiempo completo con la aplicación SelectorGadget. Se puede ver el código CSS utilizado en R.

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 carácter 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 Alejandria 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	Rodrigo 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 imparten 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

Definimos el tamaño de un grupo como el número de alumnos que tiene cada clase. Hicimos una función en *R* que simula el tamaño de los grupos con respecto a los profesores. 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).

La función realiza los siguientes pasos (para cada uno de los profesores):

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. Obtener un número con la función `sample(a:b, size=1)`.
4. Regresar el número obtenido en el paso anterior.

4.5. Simulación de solicitudes de profesores

La asignación de horarios depende de los grupos que soliciten los profesores. A partir de un esqueleto de horarios, los profesores eligen las materias y horas en las que quieren dar clase. Dado que no contamos con esa información decidimos simular la elección de materias y horarios basados en información 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: Nos ayuda para la generación de los esqueletos pero 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 dicho 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 por semestres y por hora. El procedimiento que seguimos es 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.

	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.

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.

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_demandas_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 exemplificar 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_demandas_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 generar dicha distribución con la función *normalmixEM()*. La función regresa un objeto de tipo *mixEM* el cual contiene un 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 y se define una distribución *a priori*. Con ellos se obtienen los parámetros de una distribución *a posteriori*. Éstos se convierten en los parámetros iniciales para la siguiente iteración. Este proceso lo realiza la función `normalmixEM()` hasta encontrar el modelo de mezcla de normales.

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.

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.

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_demandas_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_demandas_alumnos*. 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 y 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 tiene más información y los valores iniciales son mejores. Se puede ver el pico de las 10hrs, también se observa que se toman en cuenta los picos de las 14hrs, 16hrs y 18hrs.

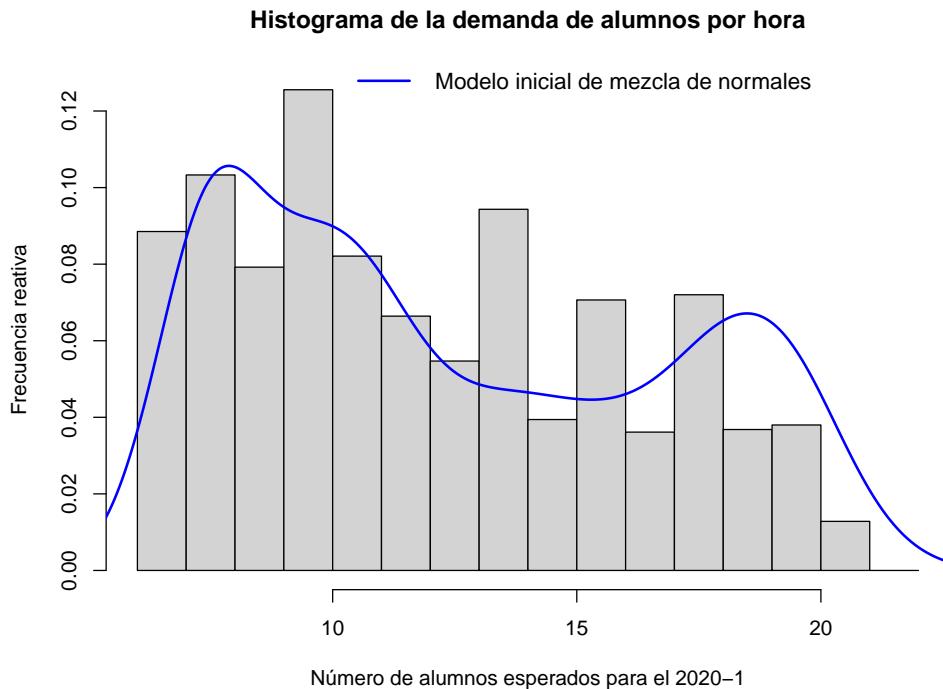


Figura 4.15: Se muestra el histograma del número de alumnos esperados por hora. La línea azul corresponde a la densidad ajustada de 1000 números aleatorios con la distribución obtenida con el modelo inicial de mezcla de normales.

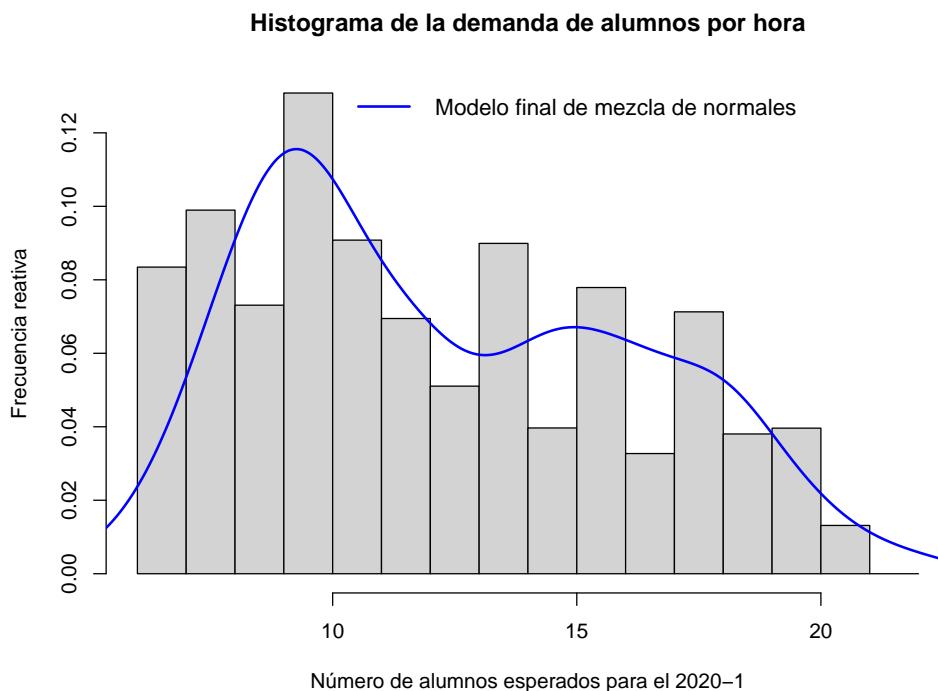


Figura 4.16: Se muestra el histograma del número de alumnos esperados por hora. La línea azul corresponde a la densidad ajustada de 1000 números aleatorios con la distribución obtenida con el modelo final de mezcla de normales.

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 calificaciones 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 $\frac{D_{h,j}^0 - D'_{h,j}}{D_{h,j}^0}$.
3. Si $D_{h,j}^0 = 0$ entonces $C_{h,j} = 1$ si faltan alumnos y $C_{h,j} = -1$ si sobran alumnos, es decir:

$$C_{h,j} = \begin{cases} 1 & \text{si } D_{h,j}^0 > D'_{h,j} \\ -1 & \text{si } D_{h,j}^0 < D'_{h,j} \\ 0 & \text{e.o.c.} \end{cases}$$
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 una. 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.

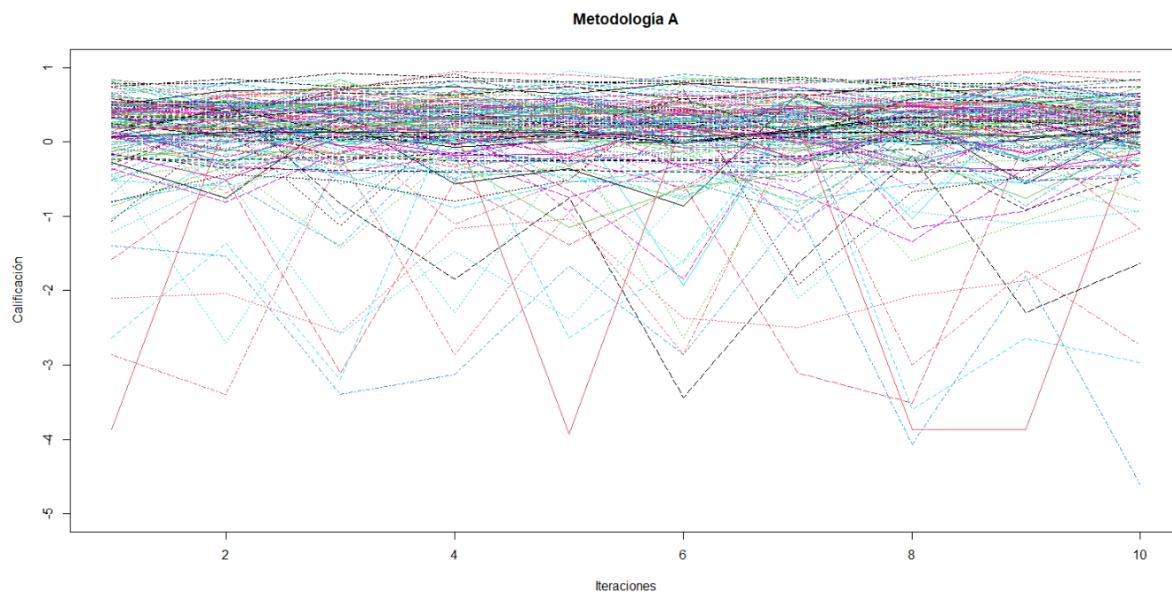


Figura 4.17: Se muestran las calificaciones por materia de la metodología A.

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.

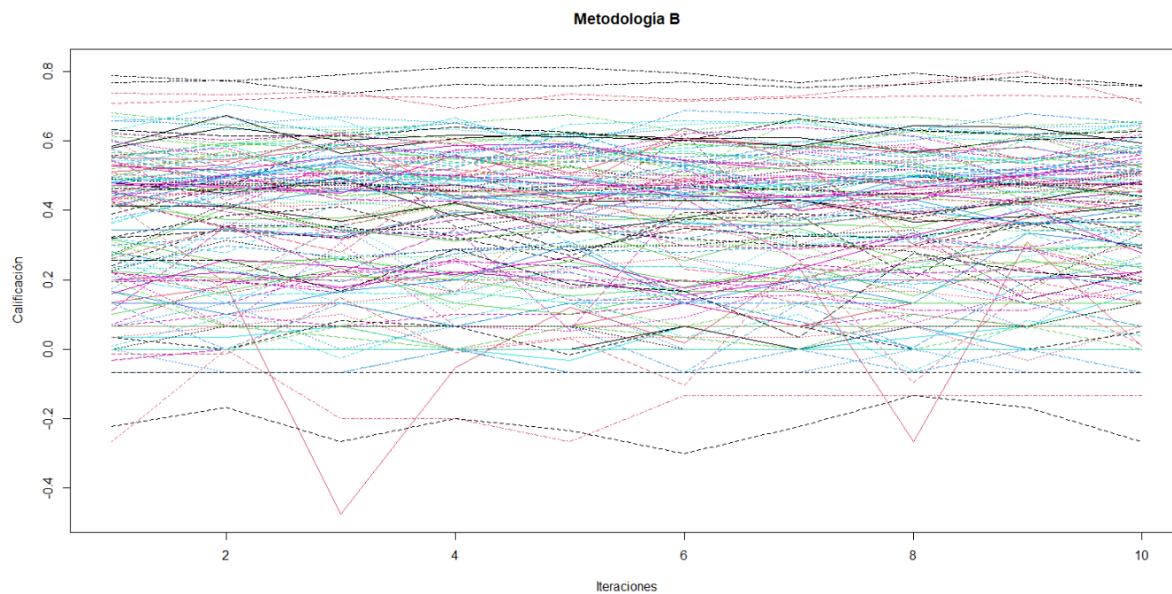


Figura 4.18: Se muestran las calificaciones por materia de la metodología B.

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.

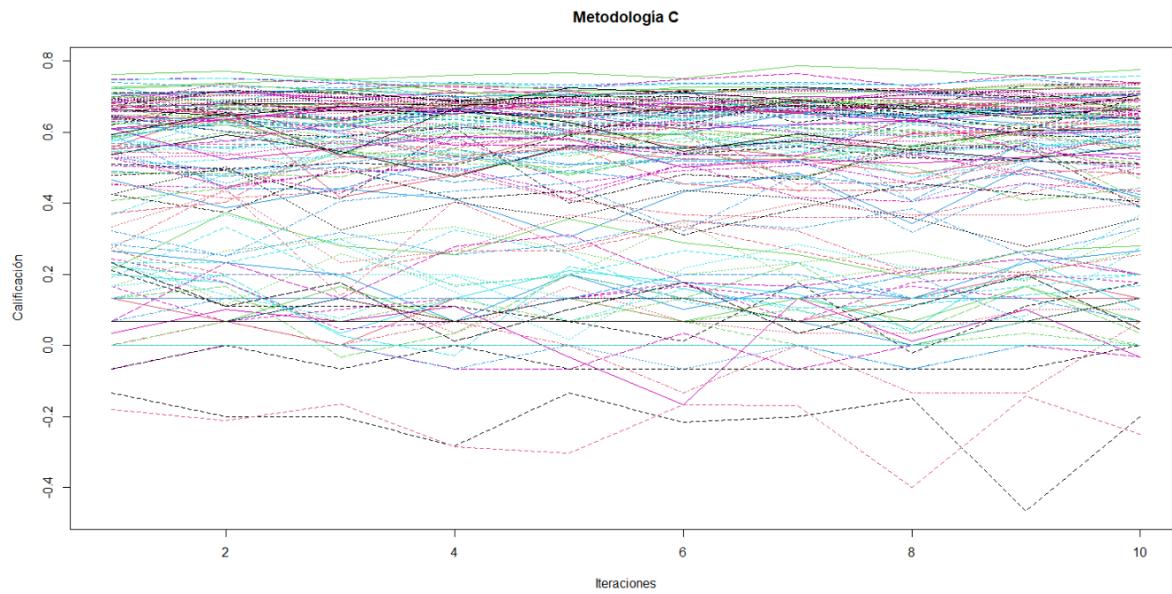


Figura 4.19: Se muestran las calificaciones por materia de la metodología C.

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 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.

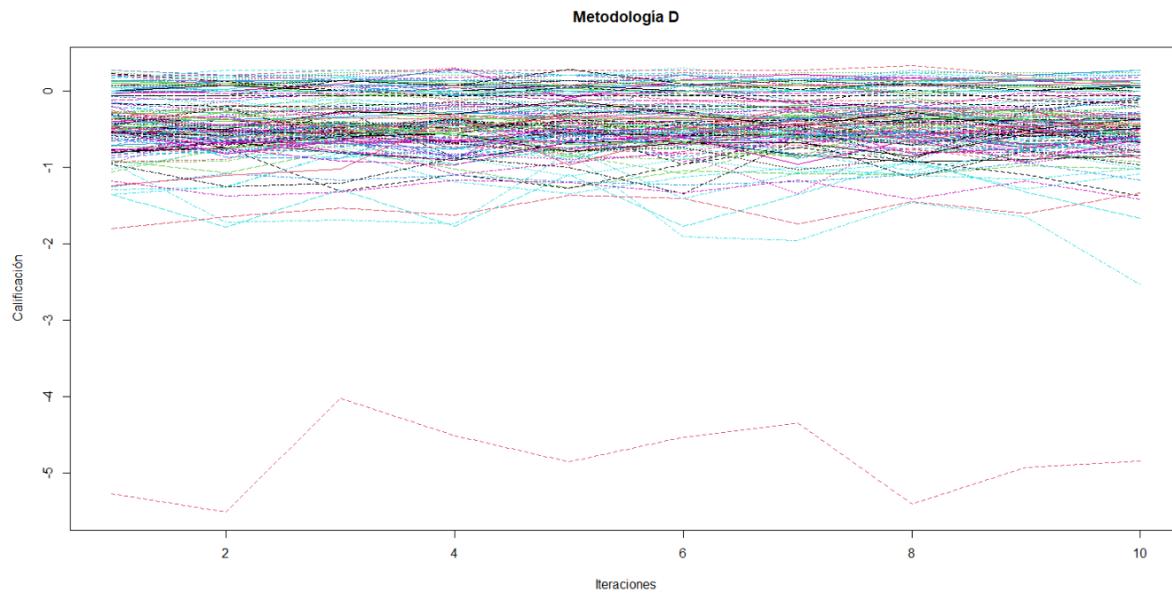


Figura 4.20: Se muestran las calificaciones por materia de la metodología D.

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.

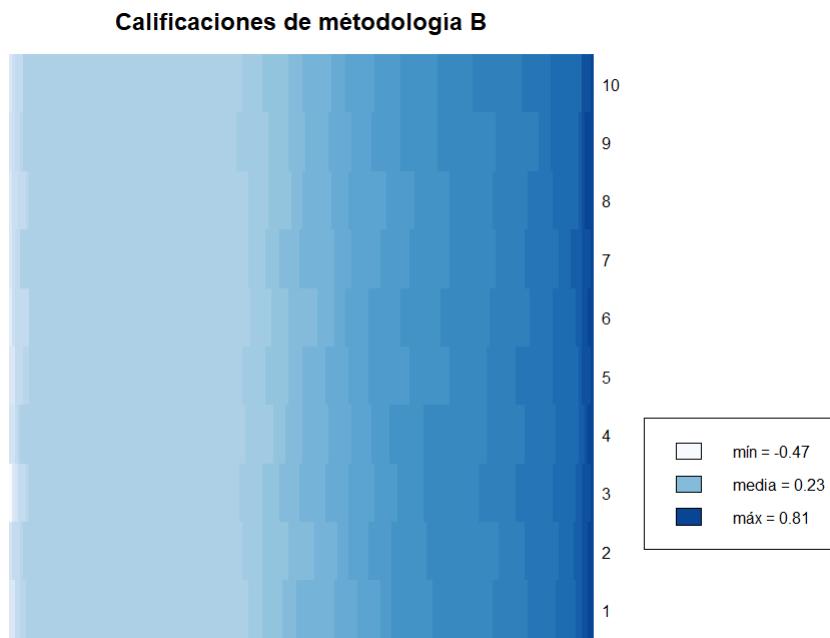


Figura 4.21: Se muestra el heatmap de las calificaciones por materia de la metodología B.

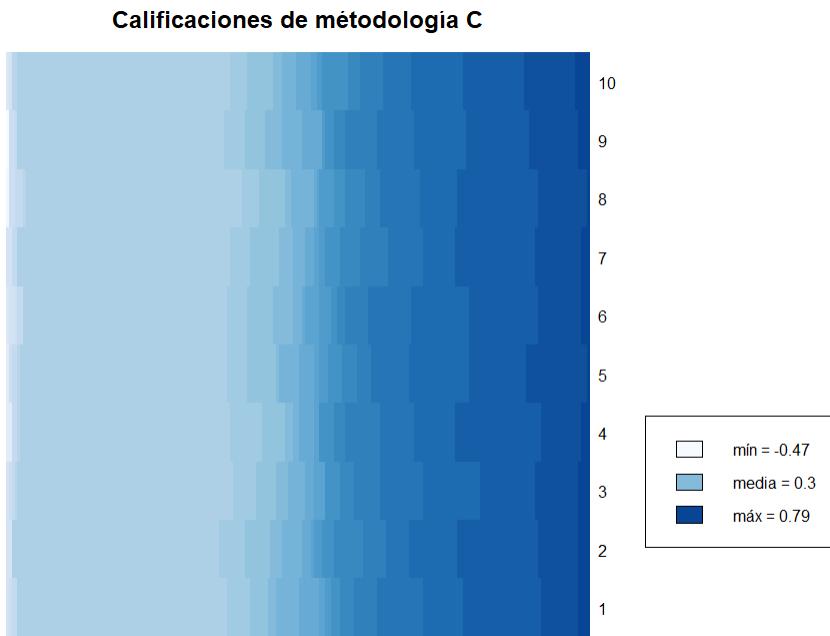


Figura 4.22: Se muestra el heatmap de las calificaciones por materia de la metodología C.

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_demandas_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

Para simular la matriz *mat_esqueleto*, con el modelo de mezcla de normales, necesitamos simular la información de varios esqueletos. Para ello vamos a considerar 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_esqueleto*

La matriz *mat_esqueleto* 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 . Dicha matriz depende de la demanda de alumnos y de las solicitudes de los profesores.

Los pasos que seguimos para obtener la matriz 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_demandas_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).
```

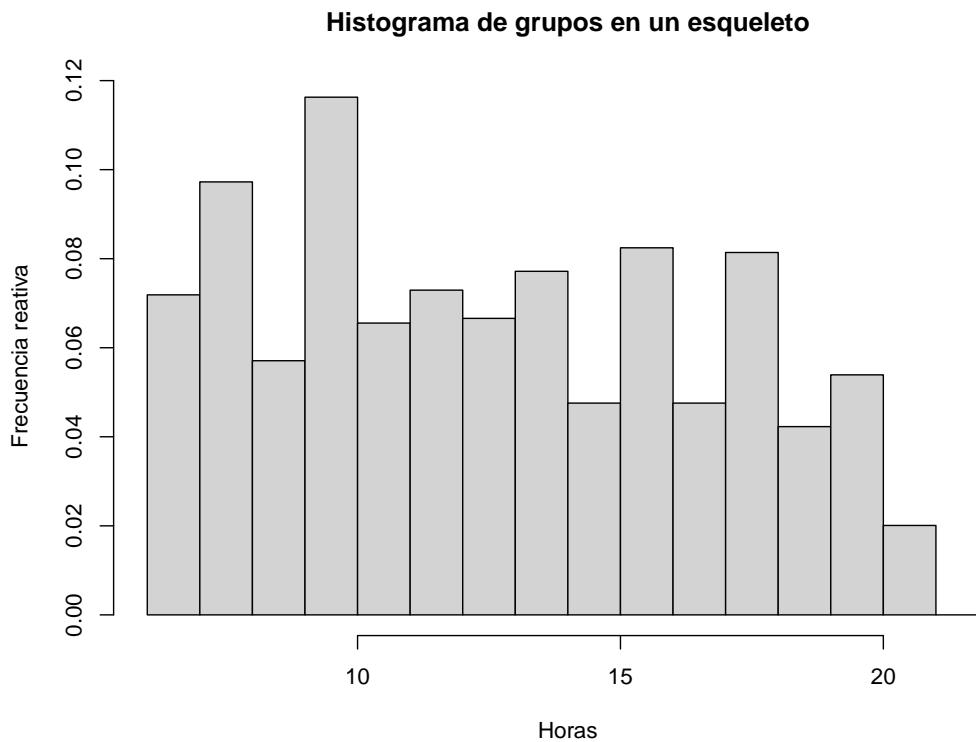


Figura 4.23: Se muestra el histograma con el número de grupos por hora en el esqueleto inicial.

Pasos a repetir ($n = 2, \dots, n_{rep}$):

1. Obtener D'_n con la función *gen_mat_demandas_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))`

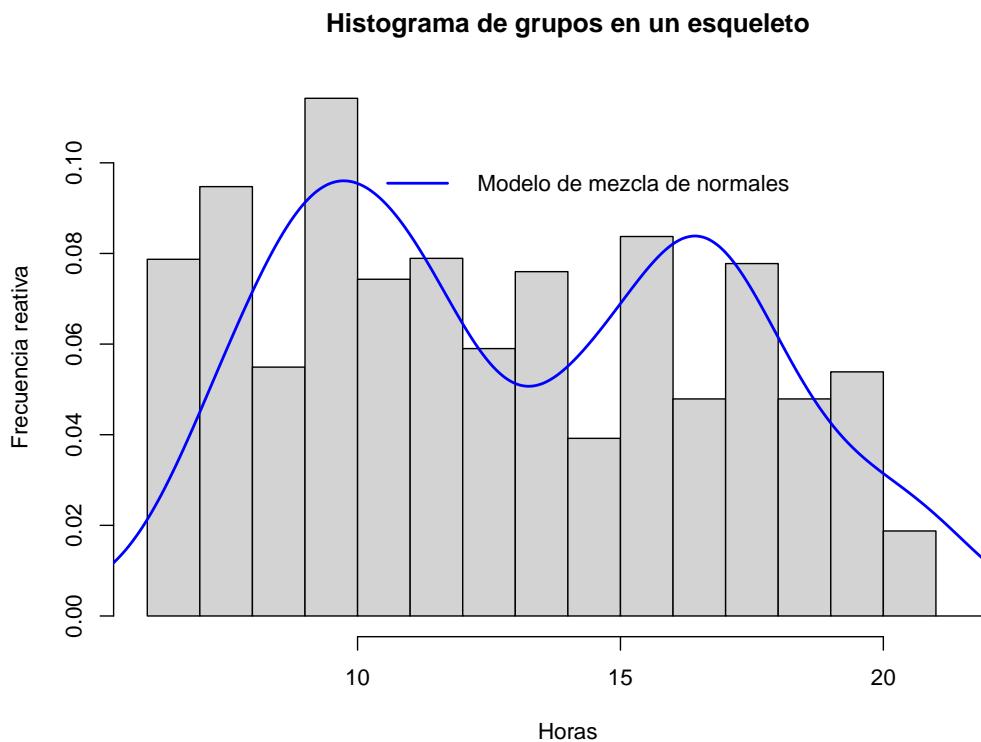


Figura 4.24: Se muestra el histograma con el número de grupos por hora de todos los esqueletos simulados. La línea azul representa la distribución ajustada por el modelo final.

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
8-9	2	3	0	0	1	2	4
9-10	3	0	0	0	1	1	3
10-11	0	2	0	0	2	2	4
11-12	0	0	1	0	2	2	1
12-13	1	1	1	0	2	0	3
13-14	0	1	0	0	2	0	0
14-15	0	0	0	0	1	1	0
15-16	1	0	0	0	0	1	2
16-17	0	1	0	0	0	0	2
17-18	0	0	0	0	1	0	2
18-19	1	1	0	0	1	0	3
19-20	0	1	1	0	0	0	0
20-21	0	0	0	0	0	2	0
21-22	1	0	0	0	1	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.

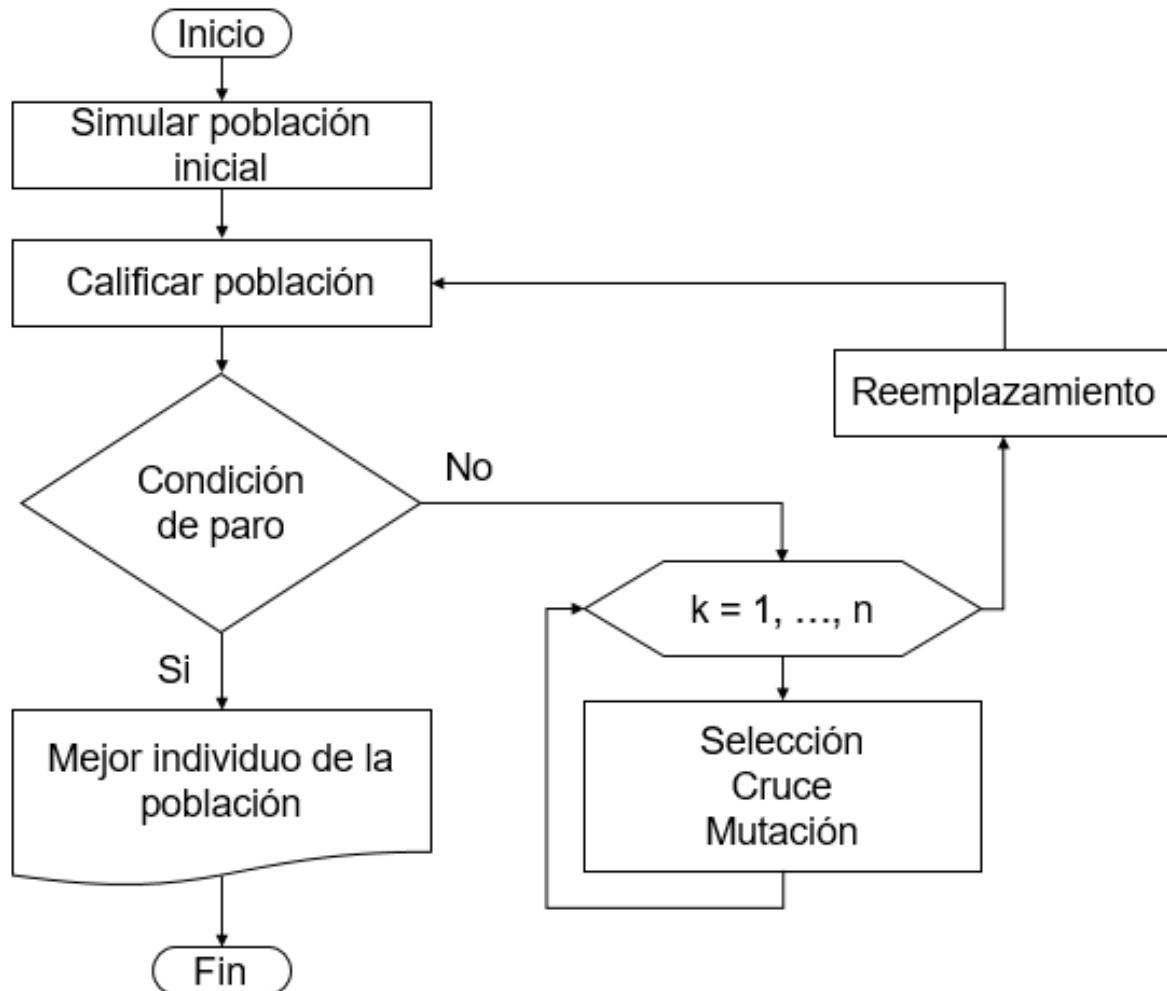


Figura 5.1: Se muestra el diagrama de flujo que se sigue en el Algoritmo Genético.

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 tiene una generación cuando se han creado n hijos y se puede reemplazar la población.

Los pasos que seguimos para obtener la asignación final 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).

Para cada asignación, comparamos el número de solicitudes con el número de grupos simulados en el esqueleto. Ésto por materia y por hora. Se tienen dos casos:

- a) El número de solicitudes es mayor al número de grupos simulados. En este caso se asignan los grupos tomados de una muestra aleatoria de las solicitudes. El tamaño de la muestra es igual al número de grupos simulados para la materia y hora correspondientes.
- b) El número de solicitudes es menor o igual al número de grupos simulados. Se asignan todos los grupos en la solicitud, para la materia y hora correspondientes.

De esta manera se asignan todos los posibles grupos que se hayan solicitado. Se debe de considerar lo siguiente:

- Un profesor no puede impartir diferentes materias a la misma hora.
- Un profesor no puede impartir la misma materia en diferentes horarios.
- Un profesor no puede tener más de num_max_asig materias asignadas.

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 $\epsilon \in [0, 0.1]$.
- 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. Se quiere elegir con mayor probabilidad los genes con mejor calificación.

Las asignaciones se ordenan de menor a mayor calificación. Se desea elegir con mayor probabilidad las asignaciones con mejor calificación.

4. Elegir 2 padres

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

5. Elegir un gen

Primero se elige al azar un parente (ambos tienen probabilidad 0.5). Una vez que se eligió un parente, seleccionar un gen: $\mathbb{P}(\text{elegir el gen } i \text{ ya ordenado}) = \frac{2i}{g(g+1)}$, 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 la probabilidad de mutación, entonces el gen tiene una mutación. Si un gen muta, se elige un gen de las solicitudes pseudo-reales y se intercambia por el gen previamente seleccionado.

7. Agregar gen

Una vez seleccionado 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 parente que aún tiene información.

Con el algoritmo utilizado en la generación de la población inicial, agregar los grupos que falta por asignar de *mat_esqueleto*.

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. Ésto porque:
 - a) No hay solicitudes de esas materias a esas horas.
 - b) Se asignaron otras materias a los profesores en esas horas.
- Es posible que la asignación final no sea la mejor calificada de todas las generaciones.

5.2. Resultados del Algoritmo Genético

Hicimos algunas pruebas con el AG para definir los parámetros de la asignación final. En la Tabla 5.1 se puede ver un resumen con los resultados de dichas pruebas. En la primer columna vemos el número de generaciones, para cada prueba. En la segunda columna se encuentra el tamaño de la población, para cada generación. En la tercera columna se observa la calificación del mejor elemento de todas las generaciones. En la cuarta columna se muestra el número de genes de la asignación final. Finalmente en la quinta columna se puede ver la calificación de la asignación final de cada prueba.

En las últimas 2 pruebas la asignación final no tuvo la mejor calificación de todas las asignaciones generadas. Ésto puede ocurrir porque el AG explora diferentes posibilidades y no necesariamente son siempre mejores que las anteriores.

Al observar la última columna notamos que al incrementar el tamaño de la población o el número de generaciones la calificación de la asignación final es cada vez mejor. Si comparamos la prueba 1 con la 2 vemos que hay una diferencia de 11.82 puntos, en la calificación. Se incrementaron 10 elementos a la población de 5 a 15 elementos. Si ahora comparamos la prueba 1 con la 3 notamos que la diferencia es de 65.24 puntos, en la calificación. En este caso se incrementó el número de generaciones de 3 a 6.

Con estas observaciones concluimos que hay un mayor incremento en la calificación cuando se aumenta el número de generaciones que el tamaño de la población.

Generaciones	Población	Mejor calif.	Genes asig. final	Calif. asig. final
3	5	-599.43	677	-599.43
3	15	-587.61	672	-587.61
6	5	-534.19	679	-534.19
6	10	-527.56	694	-527.56
8	8	-510.52	682	-510.52
10	5	-521.78	683	-530.79
10	10	-514.41	682	-515.85

Tabla 5.1: Se muestran los datos de las pruebas del Algoritmo Genético. Se observa una mejora en la calificación cuando se incrementa el número de generaciones o el tamaño de la población.

A continuación presentaremos los resultados obtenidos con el AG correspondientes a la prueba con 8 generaciones y 8 elementos por cada población. En esta prueba obtuvimos la mejor calificación en la asignación final de todas las pruebas realizadas. El número de genes varía dependiendo de cada asignación, en este caso se encuentra entre 650 y 690.

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 6 es menor al de la generación 5.

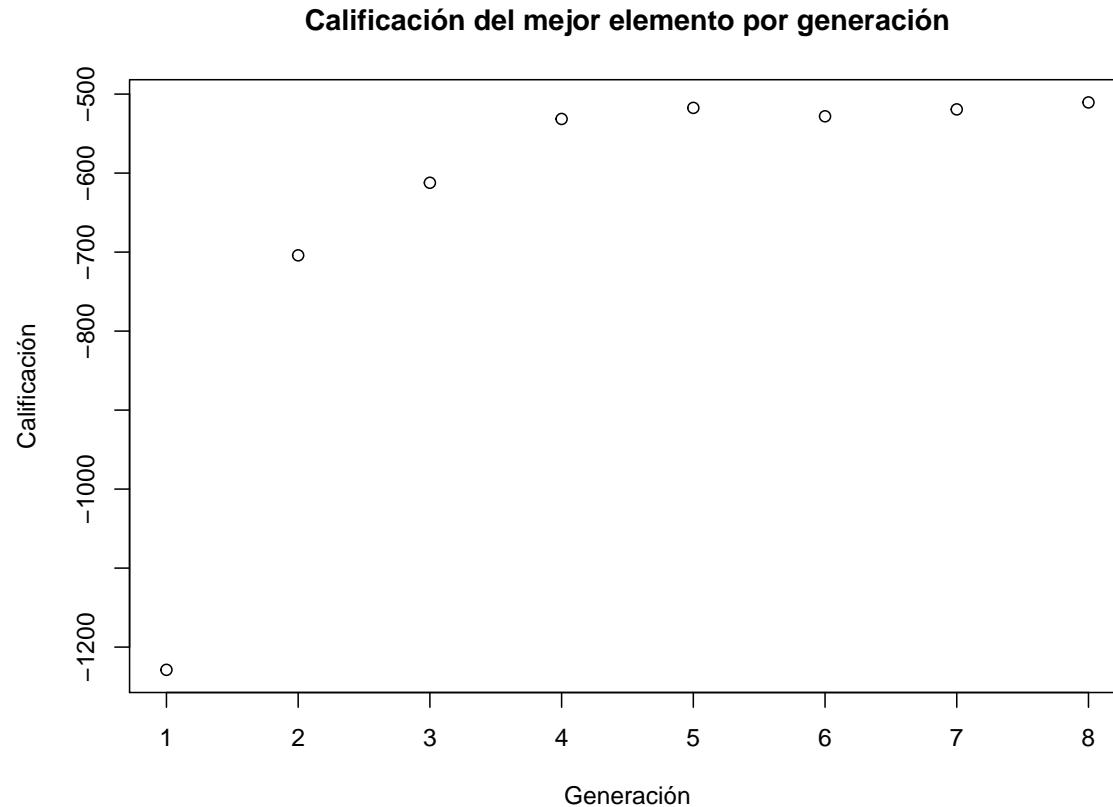


Figura 5.2: Se muestran las calificaciones de la mejor asignación por generación. Se observa una mejora considerable en la calificación de la generación 1 a la 2.

En la Figura 5.3 observamos las gráficas de caja de las calificaciones de las asignaciones, para cada generación. Hay una mejora considerable en las calificaciones de la generación 1 a 2. Los puntos representan las calificaciones de cada asignación por generación.

Notamos que las calificaciones de las asignaciones en la generación 6, son muy parecidas. A diferencia de las calificaciones en la generación 3, las cuales varían más.

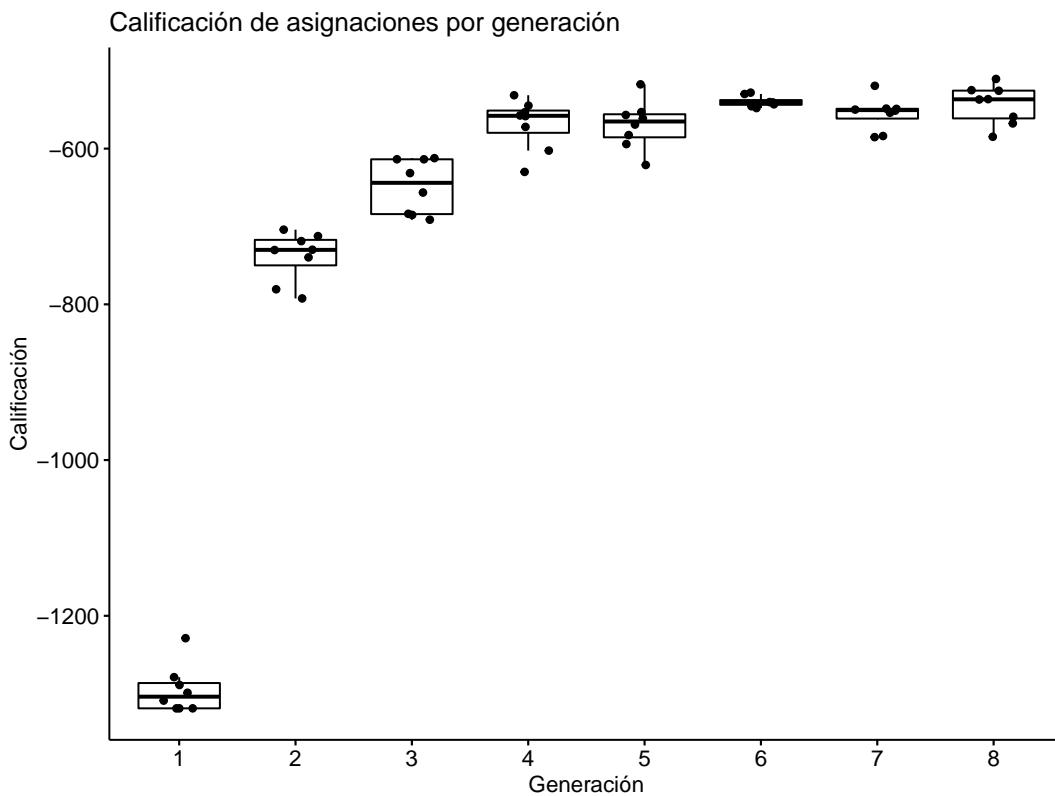


Figura 5.3: Se muestran las gráficas de caja de las calificaciones de las asignaciones por generación. Se observa una mejora considerable en la calificación de la generación 1 a la 2.

En la Figura 5.4 se puede ver el promedio de las calificaciones de las asignaciones por generación. Al igual que en las Figuras 5.2 y 5.3 se observa una mejora considerable en la calificación de la generación 1 a la 2.

Podemos observar que el promedio de las calificaciones de la generación 7 es menor que el promedio de las calificaciones de la generación 6.

En la Figura 5.5 se ve la varianza de las calificaciones de las asignaciones por generación. El valor más pequeño se observa en la generación 6. Al ver la Figura 5.3 en dicha generación, notamos que los puntos tienen valores muy similares.

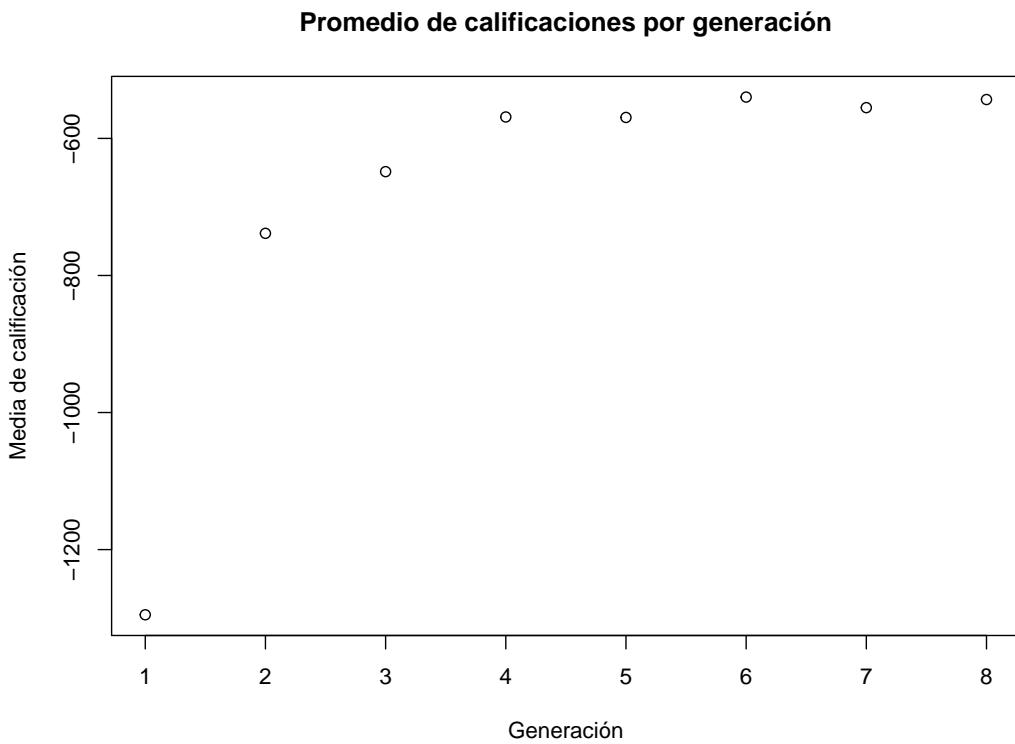


Figura 5.4: Se muestra el promedio de las calificaciones de las asignaciones por generación. Se observa una mejora considerable en la calificación de la generación 1 a la 2.

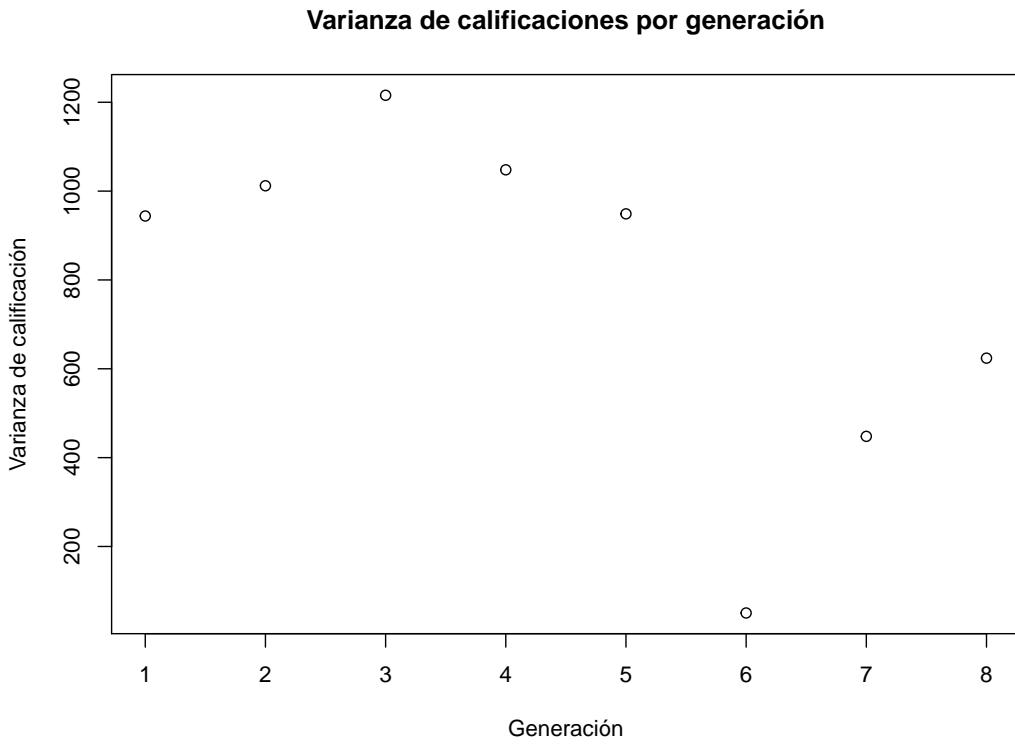


Figura 5.5: Se muestra la varianza de las calificaciones de las asignaciones por generación. Se observa una disminución considerable de la generación 3 a la 4.

En la Figura 5.6 podemos ver las gráficas de caja del número de genes en las asignaciones por generación. Se observa que los valores tienden a ser cada vez mayores. Los puntos representan el número de genes de cada asignación por generación. Notamos que la media de la generación 2 es considerablemente mayor que la media de la generación 1.

También podemos observar que el promedio del número de genes en las generaciones 7 y 8 es menor al promedio del número de genes en la generación 6. En las generaciones 4 y 7, el número de genes varía mucho.

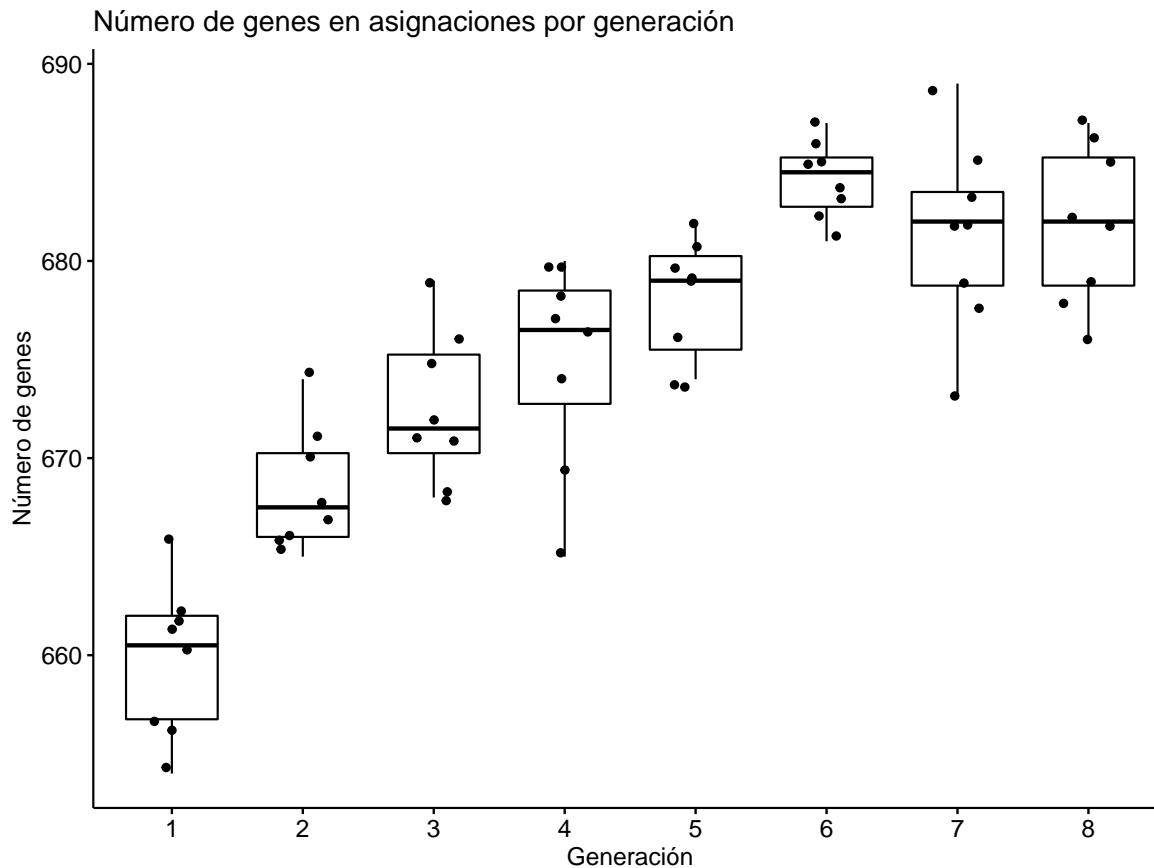


Figura 5.6: Se muestran las gráficas de caja del número de genes en asignaciones por generación. Se observa que los valores tienen una tendencia no decreciente.

Sabemos que la asignación final es el mejor elemento de la última generación. En la Tabla 5.2 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 en el Apéndice D.

Dicha matriz tiene 682 grupos asignados, cada uno con materia, profesor y horario correspondiente. En el semestre 2020-1 se tuvieron 747 grupos reales. Se simularon 1091 grupos en *mat_esqueleto*.

	Materia	Profesor	Horario
1	Administración Actuarial del Riesgo	Fernando Pérez Márquez	7
133	Cálculo Diferencial e Integral I	Elena de Oteyza de Oteyza	7
156	Cálculo Diferencial e Integral II	Héctor Méndez Lango	11
353	Inferencia Estadística	Martha Angélica Montes Fonseca	18
401	Manejo de Datos	José Alfredo Cobián Campos	13
418	Matemáticas Financieras	Irma Rocío Zavala Sierra	7
498	Modelos de Supervivencia y de Series de Tiempo	Margarita Elvira Chávez Cano	10
505	Modelos no Paramétricos y de Regresión	Jaime Vázquez Alamilla	8
510	Modelos no Paramétricos y de Regresión	Lizbeth Naranjo Albarrán	11
534	Probabilidad I	María Asunción Begoña Fernández Fernández	10
539	Probabilidad II	Daniel Alejandro Zurita Gutiérrez	8
561	Procesos Estocásticos I	Arrigo Coen Coria	10
567	Productos Financieros Derivados	Jesús Agustín Cano Garcés	10
679	Variable Compleja I	Micho Durdevich Lucic	16
682	Variable Compleja I	Jesús Manuel Mayorquín García	19

Tabla 5.2: Se muestra una submatriz de la asignación final. Cada renglón tiene la información de un grupo con una materia, profesor y horario asignado.

A continuación presentamos algunos porcentajes que nos permiten hacer diferentes comparaciones. Las materias optativas y obligatorias que se mencionan, se consideran con respecto a la carrera de Actuaría.

$$\begin{aligned}
 \frac{682}{1091} &= 62.51\% \text{ grupos asignados} \\
 \frac{241 + 18}{1091} &= 23.73\% \text{ grupos de optativas e inglés sin asignación} \\
 \frac{150}{1091} &= 13.74\% \text{ grupos de obligatorias sin asignación} \\
 \frac{682}{747} &= 91.29\% \text{ grupos asignados comparados con el 2020-1}
 \end{aligned}$$

Notamos que casi una cuarta parte de los grupos simulados en *mat_esqueleto* corresponde a materias optativas y de inglés. Ésto ocurre por dos principales motivos:

1. No hay solicitudes en *mat_solicitudes* de los grupos no asignados. Existen materias optativas que no se imparten todos los semestres. Por ello hay ocasiones en las que no se simula la solicitud.
2. Se les asigna otra materia a esa hora a los profesores que hayan solicitado alguna optativa. Usualmente se les asigna una obligatoria a la hora en la que solicitaron la optativa.

La principal razón por la que se tiene alrededor de un 10% de materias obligatorias sin asignar es por la forma en la que definimos *mat_esqueleto*. El modelo de mezcla de normales nos permite que los grupos se distribuyan a lo largo de las horas sin tener picos demasiado altos. Ésto se basa en el supuesto de que los alumnos tienen la misma preferencia por una materia a cierta hora, o a la hora anterior o a la siguiente. Es decir, si un alumno elige *Probabilidad I* a las 10hrs, le es indistinto si la toma a las 9hrs, a las 10hrs o a las 11hrs.

Este fenómeno se observa con mayor claridad en las materias con horarios muy específicos como *Teoría del Seguro*, *Matemáticas Actuariales para Seguro de Daños*, *Fianzas y Reaseguro* o *Matemáticas Actuariales del Seguro de Personas I*. En el caso de *Teoría del Seguro* se simularon 4 grupos en *mat_esqueleto*, a las 8, 10, 13 y 19hrs. Las solicitudes simuladas se pueden ver en la Figura 5.7. Notamos que sólo hay solicitudes a las 8 y a las 19hrs. Con estos datos el algoritmo sólo puede asignar dos grupos, uno a las 8hrs y el otro a las 19hrs.

Profesor	TC	Materia	Num_Materia	Horario
Agustín Peralta Cuellar	0	Teoría del Seguro	59	8
Alfonso Parrao Guzman	0	Teoría del Seguro	59	19
Alfonso Parrao Guzman	0	Teoría del Seguro	59	8
Arturo Roldán López	0	Teoría del Seguro	59	19
David López Hernández	0	Teoría del Seguro	59	8
Erik Alexander Castro Loyo	0	Teoría del Seguro	59	19
Jose Gabriel Cardenas Salmeron	0	Teoría del Seguro	59	8
Karen Lanzguerrero Obeid	0	Teoría del Seguro	59	8
Karina Vargas Cruz	0	Teoría del Seguro	59	8
María Patricia Luna Díaz	0	Teoría del Seguro	59	8
Mauricio Gabriel Arredondo Fernández Cano	0	Teoría del Seguro	59	8
Pedro Aguilar Beltrán	0	Teoría del Seguro	59	8
Ricardo Villegas Azcorra	0	Teoría del Seguro	59	8
Samadhi Cervantes Torres	0	Teoría del Seguro	59	8

Figura 5.7: Se muestran las solicitudes simuladas de la materia “Teoría del Seguro” para el semestre 2020-2.

Por estas razones el algoritmo asigna alrededor del 60% de los grupos simulados. Para aumentar este porcentaje realizamos el siguiente procedimiento:

1. Simular *mat_esqueleto* (ver Sección 4.10).
2. Simular solicitudes pseudo-reales (ver Sección 4.5).
3. Crear un archivo *xlsx* con 3 hojas:
 - Horario: Contiene una asignación predefinida con los grupos que se desean tener en la asignación final.
 - Materias: Contiene el nombre de las materias del vector *vec_nom_materias_total* (ver Sección 4.1).

- Profesores: Contiene la matriz *mat_nom_prof_total*, con la información de cada profesor (ver Sección 4.3).
4. Leer la hoja *Horario* del archivo creado con la función `read_excel`, en *R*.
 5. Eliminar la información de las solicitudes correspondiente a la asignación predefinida. Se eliminan los grupos a esa hora o con esa materia.
 6. Eliminar los grupos de la asignación predefinida en *mat_esqueleto*.
 7. Realizar el AG con las nuevas matrices de esqueleto y de solicitudes.
 8. Agregar los grupos de la asignación predefinida a los grupos de la asignación obtenida con el AG.

De esta manera se pueden definir manualmente grupos difíciles de asignar con el AG. Pueden ser materias optativas u obligatorias. Dado que se elimina la información en las solicitudes antes de realizar el AG, no se tienen grupos repetidos en la última asignación.

Cabe aclarar que en el archivo *xlsx*, las hojas *Materias* y *Profesores* nos sirven de guía para llenar de manera adecuada la hoja *Horario*. Evitando los problemas vistos en las Secciones 4.1 y 4.3 de falta de acentos, nombres incompletos, entre otros.

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.
- El algoritmo propuesto puede ser aplicado en otras escuelas o facultades que cuenten con características similares a las de la Facultad. También puede ser utilizado para resolver problemas generales de asignación de recursos.
- En particular, para las carreras de la Facultad que no pertenecen al Departamento de Matemáticas, el programa realizado se puede ajustar. Se debe de considerar lo siguiente, (cada inciso tiene un ejemplo):
 - a) Las materias impartidas en los laboratorios duran 2 o más horas: <http://www.fciencias.unam.mx/docencia/horarios/20191/1081/830>.
 - b) No todas las materias se imparten todos los días: <http://www.fciencias.unam.mx/docencia/horarios/20192/1439/1419>.
 - c) Existen materias que no duran horas enteras: <http://www.fciencias.unam.mx/docencia/horarios/20192/1081/104>.
 - d) Hay materias que no inician en horas enteras: <http://www.fciencias.unam.mx/docencia/horarios/20172/181/1601>.

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)** (297) -> 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)** (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, \dots, X_n una muestra aleatoria de una población con función de densidad de probabilidad Poisson(λ). Su función de densidad es:

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

$$\begin{aligned}\mathcal{L}(X_1, X_2, \dots, X_n; \lambda) &= \prod_{i=1}^n \left(e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \right) \\ &= e^{-n\lambda} \frac{\sum_{i=1}^n x_i}{\prod_{i=1}^n x_i!}\end{aligned}$$

Sacamos ln

$$\ln \mathcal{L}(X_1, X_2, \dots, X_n; \lambda) = -n\lambda + \sum_{i=1}^n x_i \ln \lambda - \ln \prod_{i=1}^n x_i!$$

Derivamos con respecto a λ

$$\frac{\partial}{\partial \lambda} \ln \mathcal{L}(\underline{X}; \lambda) = -n + \frac{\sum_{i=1}^n x_i}{\lambda}$$

Igualamos a cero

$$\begin{aligned}-n + \frac{\sum_{i=1}^n x_i}{\lambda} &= 0 \\ \Rightarrow \quad &\end{aligned}$$

Despejamos λ

$$\hat{\lambda} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}$$

Derivamos otra vez

$$\begin{aligned}\frac{\partial^2}{\partial \lambda^2} \ln \mathcal{L}(\underline{X}; \lambda) &= -\frac{\sum_{i=1}^n x_i}{\lambda^2} < 0 \\ \therefore \hat{\lambda} &= \bar{x} \text{ es el estimador máximo verosímil}\end{aligned}$$

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 630 grupos asignados, cada uno con materia, profesor y horario correspondiente.

	Materia	Profesor	Horario
1	Administración Actuarial del Riesgo	Fernando Pérez Márquez	7
2	Administración Actuarial del Riesgo	Samadhi Cervantes Torres	8
3	Administración de Riesgos Financieros	María Araceli Bernabe Rocha	8
4	Administración Financiera	Pedro Luis Soto Tejeda	8
5	Álgebra Lineal I	Irvin Arellano Rosas	7
6	Álgebra Lineal I	León Felipe Villalobos Sánchez	8
7	Álgebra Lineal I	Rodrigo Edmundo Cepeda Morales	8
8	Álgebra Lineal I	Andres Barei Bueno	8
9	Álgebra Lineal I	César Alejandro Rincón Orta	8
10	Álgebra Lineal I	Karina García Buendía	8
11	Álgebra Lineal I	Alma Violeta García López	8
12	Álgebra Lineal I	Francisco Marmolejo Rivas	9
13	Álgebra Lineal I	Gabriela Campero Arena	9
14	Álgebra Lineal I	Hugo Arizmendi Peimbert	10
15	Álgebra Lineal I	Jorge Luis Arocha Pérez	10
16	Álgebra Lineal I	Adolfo Guillot Santiago	10
17	Álgebra Lineal I	Gustavo Amilcar Saldaña Moncada	10
18	Álgebra Lineal I	María del Carmen Heréndira Gómez Laveaga	12
19	Álgebra Lineal I	Adriana Léon Montes	13
20	Álgebra Lineal I	José Pozo Martínez	14

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
21	Álgebra Lineal I	Juan Orendain Almada	15
22	Álgebra Lineal I	Iván Fernando Vilchis Montalvo	17
23	Álgebra Lineal I	José Cruz García Zagal	18
24	Álgebra Lineal I	Jorge Andrés Rosas Ávila	18
25	Álgebra Lineal I	Víctor Manuel Ávila Baez	18
26	Álgebra Lineal I	Israel Zamorano Romero	18
27	Álgebra Lineal I	José Miguel Salazar Montiel	19
28	Álgebra Lineal I	Raúl Bartolo Martínez	20
29	Álgebra Lineal II	Daniela Mariyet Terán Guerrero	7
30	Álgebra Lineal II	Benjamín Iván Juárez Santos	7
31	Álgebra Lineal II	Carlos Alberto Serrato Hernández	8
32	Álgebra Lineal II	Gustavo Amilcar Saldaña Moncada	9
33	Álgebra Lineal II	Juan Carlos García Altamirano	10
34	Álgebra Lineal II	Rocío del Pilar Aguilar Benítez	10
35	Álgebra Lineal II	Guillermo Javier Francisco Sienra Loera	10
36	Álgebra Lineal II	Juan Morales Rodríguez	10
37	Álgebra Lineal II	Jorge Luis Arocha Pérez	13
38	Álgebra Lineal II	Alejandro Alvarado García	14
39	Álgebra Lineal II	Lorena Morales Callejas	16
40	Álgebra Lineal II	William José Gallardo	17
41	Álgebra Lineal II	José Miguel Salazar Montiel	18
42	Álgebra Moderna I	Mindy Yaneli Huerta Pérez	7
43	Álgebra Moderna I	José Cruz García Zagal	15
44	Álgebra Moderna I	Sergio Hiroki Koike Quintanar	15
45	Álgebra Moderna I	José Collins Castro	16
46	Álgebra Moderna I	Christopher Román Silva Sarabia	18
47	Álgebra Moderna I	Edgar Migueles Pérez	19
48	Álgebra Moderna II	José Gabriel Ocampo Márquez	7
49	Álgebra Moderna II	Diana Avella Alaminos	10
50	Álgebra Moderna II	Bertha María Tomé Arreola	10
51	Álgebra Moderna II	Mauricio Gabriel Medina Bárcenas	14

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
52	Álgebra Moderna II	Eugenia O'Reilly Regueiro	18
53	Álgebra Moderna III	Frank Patrick Murphy Hernández	7
54	Álgebra Moderna III	Valente Santiago Vargas	9
55	Álgebra Moderna III	Marco Federico Larrea Schiavon	17
56	Álgebra Superior I	Alejandra Osiris Romero Juárez	7
57	Álgebra Superior I	Lorena Morales Callejas	8
58	Álgebra Superior I	Daniela Mariyet Terán Guerrero	8
59	Álgebra Superior I	Ernesto Mayorga Saucedo	8
60	Álgebra Superior I	Lilia Guadalupe Sánchez Terán	8
61	Álgebra Superior I	Martha Takane Imay	10
62	Álgebra Superior I	Francisco Larrión Riveroll	10
63	Álgebra Superior I	Rita Esther Zuazua Vega	10
64	Álgebra Superior I	Alejandro Bravo Mojica	10
65	Álgebra Superior I	Daniel Labardini Fragoso	14
66	Álgebra Superior I	Paloma Hernández Zapata	14
67	Álgebra Superior I	Marcelino Perello Valls	16
68	Álgebra Superior I	Pavel Ramos Martínez	17
69	Álgebra Superior I	Rodrigo Domínguez López	17
70	Álgebra Superior I	Marco Antonio Acevedo Cardona	18
71	Álgebra Superior I	Jaime Castro Pérez	18
72	Álgebra Superior I	Natalia Bárbara Mantilla Beniers	18
73	Álgebra Superior I	Hugo Rincón Galeana	19
74	Álgebra Superior II	Tania Eréndira Rivera Torres	7
75	Álgebra Superior II	Rolando Gómez Macedo	7
76	Álgebra Superior II	Gabriela Cervantes Piza	8
77	Álgebra Superior II	César Alejandro Rincón Orta	9
78	Álgebra Superior II	Carmen Martínez Adame Isaías	9
79	Álgebra Superior II	Gerardo Camacho de la Rosa	10
80	Álgebra Superior II	Araceli Guzmán Tristán	10
81	Álgebra Superior II	Juan Morales Jr Rodríguez	11
82	Álgebra Superior II	Juan Morales Rodríguez	12
83	Álgebra Superior II	Antonio Lascurain Orive	12
84	Álgebra Superior II	José Patricio Sánchez Hernández	14

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
85	Álgebra Superior II	Alejandro Dorantes Aldama	15
86	Álgebra Superior II	Jorge Enrique Vega Acevedo	15
87	Álgebra Superior II	Emiliano Geneyro Squarzon	16
88	Álgebra Superior II	Ana Patricia Kuri González	16
89	Álgebra Superior II	Sebastián Pardo Guerra	17
90	Álgebra Superior II	Leonardo Faustinos Morales	18
91	Álgebra Superior II	Joseph Christian Sierra Gutiérrez	20
92	Análisis del México Contemporáneo	Luis Alberto Galván Contreras	8
93	Análisis del México Contemporáneo	Silvia Alonso Reyes	10
94	Análisis del México Contemporáneo	Gustavo Esteban Andrade Díaz	17
95	Análisis Matemático I	Ricardo Alberto Weder Zaninovich	7
96	Análisis Matemático I	Rodrigo Jesús Hernández Gutiérrez	7
97	Análisis Matemático I	Roxana Wendoline Ruiz Aguilar	8
98	Análisis Matemático I	Roberto Méndez Rosas	10
99	Análisis Matemático I	Agustín Ontiveros Pineda	10
100	Análisis Matemático I	Carmen Martínez Adame Isaías	10
101	Análisis Matemático I	Sergey Antonyan	10
102	Análisis Matemático I	José Lino Samaniego Mendoza	11
103	Análisis Matemático I	Alberto León Kushner Schnur	13
104	Análisis Matemático I	Mario Eudave Muñoz	14
105	Análisis Matemático I	César Eduardo Sousa Mondragón	14
106	Análisis Matemático I	Grissel Santiago González	15
107	Análisis Matemático I	Vinicio Antonio Gómez Gutiérrez	16
108	Análisis Matemático I	Roberto Pichardo Mendoza	16
109	Análisis Matemático I	Manuel Hernández Hernández	17
110	Análisis Matemático I	Alfredo Reyes Vázquez	18
111	Análisis Matemático II	Esteban Librado Hernández Escamilla	7
112	Análisis Matemático II	Mónica Alicia Clapp Jiménez-Labora	10
113	Análisis Matemático II	Angel Manuel Carrillo Hoyo	10

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
114	Análisis Matemático II	Manuel Jesús Falconi Magaña	10
115	Análisis Matemático II	Ana Meda Guardiola	10
116	Análisis Matemático II	Gerardo Sánchez Licea	13
117	Análisis Matemático II	Raybel Andrés García Ancona	14
118	Análisis Matemático II	Nils Heye Ackermann	18
119	Análisis Matemático II	Manuel Hernández Hernández	18
120	Análisis Matemático II	Juan Manuel de la Huerta Jiménez	19
121	Análisis Matemático III	Magali Louise Marie Folch Gabayet	10
122	Análisis Matemático III	Bernardo Vargas Cárdenas	12
123	Análisis Matemático III	Guillermo Grabinsky Steider	16
124	Análisis Multivariado	Guadalupe Eunice Campiran García	13
125	Análisis Numérico	Antonio Carrillo Ledesma	7
126	Análisis Numérico	Pablo Barrera Sánchez	10
127	Análisis Numérico	Úrsula Xiomara Iturrarán Vi-veros	10
128	Análisis Numérico	Miriam Sosa Díaz	14
129	Análisis Numérico	Mario Medina Torres	17
130	Análisis Numérico	Miguel Ángel Pérez León	19
131	Análisis Numérico	Agustín Alberto Rosas Medina	19
132	Autómatas y Lenguajes Formales	Elisa Viso Gurovich	10
133	Cálculo Diferencial e Integral I	Elena de Oteyza de Oteyza	7
134	Cálculo Diferencial e Integral I	Joel García León	10
135	Cálculo Diferencial e Integral I	Antonmaria Gerolamo Enrico Minzoni Alessio	10
136	Cálculo Diferencial e Integral I	Alessio Franci	10
137	Cálculo Diferencial e Integral I	Javier Fernández García	12
138	Cálculo Diferencial e Integral I	María de Lourdes Esteva Peralta	12
139	Cálculo Diferencial e Integral I	José Antonio Gómez Ortega	12
140	Cálculo Diferencial e Integral I	Wilfrido Martínez Torres	15
141	Cálculo Diferencial e Integral I	María Lourdes Velasco Arregui	15
142	Cálculo Diferencial e Integral I	Mónica De Nova Vázquez	16
143	Cálculo Diferencial e Integral I	Julio Martín Espinosa Casares	18
144	Cálculo Diferencial e Integral I	Edgar René Hernández Martínez	20

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
145	Cálculo Diferencial e Integral I	Mario Delgadillo Torres	20
146	Cálculo Diferencial e Integral II	Esteban Rubén Hurtado Cruz	10
147	Cálculo Diferencial e Integral II	Oscar Alberto Garrido Jiménez	10
148	Cálculo Diferencial e Integral II	Carlos García Azpeitia	10
149	Cálculo Diferencial e Integral II	María del Carmen Arrillaga Arjona	10
150	Cálculo Diferencial e Integral II	Miguel Lara Aparicio	10
151	Cálculo Diferencial e Integral II	Gustavo Cruz Pacheco	10
152	Cálculo Diferencial e Integral II	Luis Manuel Hernández Gallardo	10
153	Cálculo Diferencial e Integral II	María Lourdes Velasco Arreguí	10
154	Cálculo Diferencial e Integral II	Fernando Brambila Paz	11
155	Cálculo Diferencial e Integral II	Humberto Andrés Carrillo Calvet	11
156	Cálculo Diferencial e Integral II	Héctor Méndez Lango	11
157	Cálculo Diferencial e Integral II	Leticia Contreras Sandoval	13
158	Cálculo Diferencial e Integral II	César Eduardo Sousa Mondragón	13
159	Cálculo Diferencial e Integral II	Jesús López Estrada	13
160	Cálculo Diferencial e Integral II	Héctor Bustos Castro	14
161	Cálculo Diferencial e Integral II	Roberto García Medina	15
162	Cálculo Diferencial e Integral II	David Meza Alcántara	16
163	Cálculo Diferencial e Integral II	José Santos	17
164	Cálculo Diferencial e Integral II	Francisco Javier Torres Ayala	17
165	Cálculo Diferencial e Integral II	Julio César Cedillo Sánchez	19
166	Cálculo Diferencial e Integral II	Leonel Rito Rodríguez	20
167	Cálculo Diferencial e Integral III	Felipe de Jesús Méndez Varela	7
168	Cálculo Diferencial e Integral III	Javier Fernández García	7
169	Cálculo Diferencial e Integral III	Emma Lam Osnaya	7
170	Cálculo Diferencial e Integral III	Alejandro Darío Rojas Sánchez	9
171	Cálculo Diferencial e Integral III	Adriana Vargas Quintero	10
172	Cálculo Diferencial e Integral III	Julio Martín Espinosa Casares	10
173	Cálculo Diferencial e Integral III	Pablo Alberto Lara Martínez	10
174	Cálculo Diferencial e Integral III	Omar Vigueras Herrera	11
175	Cálculo Diferencial e Integral III	Javier Páez Cárdenas	11
176	Cálculo Diferencial e Integral III	Carlos Gerardo Paniagua Ramírez	13
177	Cálculo Diferencial e Integral III	Enrique Vega Ramírez	15
178	Cálculo Diferencial e Integral III	Israel Ramos García	16

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
179	Cálculo Diferencial e Integral III	Álvaro Reyes García	18
180	Cálculo Diferencial e Integral III	Noé Francisco Verde Martínez	19
181	Cálculo Diferencial e Integral III	Alfredo Reyes Vázquez	19
182	Cálculo Diferencial e Integral IV	Sandra María Chimal Garma	7
183	Cálculo Diferencial e Integral IV	Adrián Zenteno Gutiérrez	7
184	Cálculo Diferencial e Integral IV	Abelardo Vela Ponce de León	7
185	Cálculo Diferencial e Integral IV	Alberto Manuel Aldama Garijoain	7
186	Cálculo Diferencial e Integral IV	Carlos Prieto López	10
187	Cálculo Diferencial e Integral IV	Francisco Marmolejo Rivas	10
188	Cálculo Diferencial e Integral IV	Javier Páez Cárdenas	10
189	Cálculo Diferencial e Integral IV	Héctor Méndez Lango	10
190	Cálculo Diferencial e Integral IV	Marco Antonio Montes de Oca Balderas	14
191	Cálculo Diferencial e Integral IV	Emilio Cabrera Castro	14
192	Cálculo Diferencial e Integral IV	Jefferson Edwin King Dávalos	15
193	Cálculo Diferencial e Integral IV	Héctor Jiménez Sánchez	17
194	Cálculo Diferencial e Integral IV	Oscar Alberto Garrido Jiménez	19
195	Cálculo Diferencial e Integral IV	Gabriel Gutiérrez García	19
196	Conjuntos y Lógica	César Hernández Cruz	8
197	Conjuntos y Lógica	Fernando Javier Nuñez Rosales	18
198	Contabilidad	Emilia Lorenzana Quintero	7
199	Contabilidad	María de los Ángeles Garduño Crespo	8
200	Contabilidad	Arely Palos Zepeda	17
201	Contabilidad	Elvia Sánchez Montes	21
202	Demografía	Juan Diego Terán Páez	7
203	Demografía	José Rubén Fernández Román	7
204	Demografía	Rodrigo Jiménez Uribe	7
205	Demografía	Rosalba Jasso Vargas	10
206	Demografía	Nina Castro Méndez	11
207	Demografía	Mauricio Rodríguez Abreu	17
208	Dinámica de Medios Deformables	Catalina Elizabeth Stern Forgach	8
209	Dinámica de Medios Deformables	Guillermo Monsivais Galindo	10
210	Dinámica de Medios Deformables	Consuelo García Alcántara	12
211	Dinámica de Medios Deformables	Roberto Velasco Segura	16
212	Economía	José Antonio Reyes León	7

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
213	Economía	Marco Antonio García Fernández	7
214	Economía	Luis Alejandro Aguilar Luna	8
215	Economía	Monserrat Esquivel López	10
216	Economía	Jessika Dilhery Lucas Flores	16
217	Ecuaciones Diferenciales I	Catalina Apolinar García	7
218	Ecuaciones Diferenciales I	Javier de Jesús Cortés Aguirre	7
219	Ecuaciones Diferenciales I	Maya Lol Sosa Salas	7
220	Ecuaciones Diferenciales I	Raziel Zavaleta Rodríguez	7
221	Ecuaciones Diferenciales I	Laura Ortiz Bobadilla	8
222	Ecuaciones Diferenciales I	Nicolás González Boileau	8
223	Ecuaciones Diferenciales I	Gerardo Mejía Rodríguez	10
224	Ecuaciones Diferenciales I	Luisa Márquez Rentería	10
225	Ecuaciones Diferenciales I	Araceli León Estrada	10
226	Ecuaciones Diferenciales I	Carlos Arturo Vargas Guadarrama	11
227	Ecuaciones Diferenciales I	Manuel Jesús Falconi Magaña	11
228	Ecuaciones Diferenciales I	Carlos García Azpeitia	12
229	Ecuaciones Diferenciales I	Gustavo Cruz Pacheco	13
230	Ecuaciones Diferenciales I	María del Carmen Jorge y Jorge	14
231	Ecuaciones Diferenciales I	Adrián Ulises Soto Bañuelos	17
232	Ecuaciones Diferenciales I	Karen Susana Villa Aguirre	18
233	Ecuaciones Diferenciales I	Fidencio Galicia Rodríguez	18
234	Ecuaciones Diferenciales I	Herminio Suárez Quiroz	18
235	Ecuaciones Diferenciales I	Octavio Eduardo Vizcaya Xilotl	19
236	Ecuaciones Diferenciales I	Jorge Luis Arroyo Leonor	21
237	Ecuaciones Diferenciales II	Ernesto Rosales González	8
238	Ecuaciones Diferenciales II	Catherine García Reimbert	10
239	Ecuaciones Diferenciales II	Humberto Andrés Carrillo Calvet	13
240	Ecuaciones Diferenciales II	José Luis Navarro Urrutia	16
241	Ecuaciones Diferenciales Parciales I	María de los Ángeles Sandoval Romero	11
242	Electromagnetismo I	Eugenio Ley Koo	7
243	Electromagnetismo I	Roberto Alejandro Ruelas Mayorga	7
244	Electromagnetismo I	Mirna Villavicencio Torres	8
245	Electromagnetismo I	Andrea Luisa Aburto Espina	10
246	Electromagnetismo I	Jorge Alfonso García Macedo	10
247	Electromagnetismo I	Julio Javier Martinell Benito	10

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
248	Electromagnetismo I	Alejandro Reyes Coronado	10
249	Electromagnetismo I	José Alberto Flandes Mendoza	10
250	Electromagnetismo I	Ricardo Méndez Fragoso	16
251	Electromagnetismo I	Pablo Antonio Borys Sosa	16
252	Electromagnetismo I	Antonio Castellanos Ramírez	18
253	Electromagnetismo II	Rocío Jáuregui Renaud	8
254	Electromagnetismo II	Luis Fernando Urrutia Rios	10
255	Electromagnetismo II	Ángel Prieto Ruiz	10
256	Estadística Bayesiana	Ruth Selene Fuentes García	10
257	Estadística Bayesiana	Harim García Lamont	19
258	Estructuras Discretas	Favio Ezequiel Miranda Pereira	10
259	Estructuras Discretas	Alma Rosario Arévalo Loya	14
260	Fenómenos Colectivos	Francisco Ramos Gómez	8
261	Fenómenos Colectivos	Carlos Enrique Román Velázquez	8
262	Fenómenos Colectivos	Francisco Javier Reyes Mora	8
263	Fenómenos Colectivos	Luis Felipe del Castillo Dávila	10
264	Fenómenos Colectivos	Margarita Sánchez y Sánchez	10
265	Fenómenos Colectivos	Iván Santamaría Holek	10
266	Fenómenos Colectivos	Soledad Ruiz Matus	17
267	Física Computacional	Edgar Vázquez Luis	7
268	Física Computacional	Ricardo Méndez Fragoso	7
269	Física Computacional	Enrique Cabrera Bravo	10
270	Física Computacional	Lucía Medina Gómez	10
271	Física Computacional	Roxana Mitzayé Del Castillo Vázquez	18
272	Física Estadística	Gerardo Carmona Ruiz	8
273	Física Estadística	Guillermo Ramírez Santiago	10
274	Física Estadística	Rosalío Fernando Rodríguez Zepeda	10
275	Funciones Especiales y Transformadas Integrales	Mariana Martínez González	8
276	Funciones Especiales y Transformadas Integrales	Guillermo Monsivais Galindo	8
277	Funciones Especiales y Transformadas Integrales	Eduardo Peinado Rodríguez	8
278	Funciones Especiales y Transformadas Integrales	Luis Antonio Dávalos Orozco	10
279	Funciones Especiales y Transformadas Integrales	Rosario Paredes Gutiérrez	10

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
280	Funciones Especiales y Transformadas Integrales	Mauricio García Vergara	10
281	Funciones Especiales y Transformadas Integrales	Lucía Medina Gómez	11
282	Funciones Especiales y Transformadas Integrales	Osvaldo Alfonso Téllez Nieto	13
283	Funciones Especiales y Transformadas Integrales	Ramón Gustavo Contreras Mayen	14
284	Funciones Especiales y Transformadas Integrales	Francisco Javier Martínez Farias	16
285	Geometría Analítica I	Juan Flores Torres	7
286	Geometría Analítica I	Jorge Chávez Carlos	8
287	Geometría Analítica I	Max Neumann Coto	8
288	Geometría Analítica I	Rebeca Trejo Luna	8
289	Geometría Analítica I	Fernando García Rodríguez	9
290	Geometría Analítica I	Jaqueline Rafaela Dolores Cañetas Ortega	10
291	Geometría Analítica I	Guillermo Ruiz Galván	10
292	Geometría Analítica I	Pablo Suárez Serrato	10
293	Geometría Analítica I	Jesús Ángel Núñez Zimbrón	10
294	Geometría Analítica I	Edgar Alvarez Zauco	10
295	Geometría Analítica I	Javier Bracho Carpizo	10
296	Geometría Analítica I	Oscar Alfredo Palmas Velasco	12
297	Geometría Analítica I	Guillermo Javier Francisco Sienra Loera	12
298	Geometría Analítica I	Federico Sánchez Bringas	13
299	Geometría Analítica I	Noel Jaramillo Arce	14
300	Geometría Analítica I	Leopoldo Morales López	14
301	Geometría Analítica I	Víctor Manuel Ávila Baez	17
302	Geometría Analítica I	Adrián Ulises Soto Bañuelos	18
303	Geometría Analítica I	Jorge Ortiz Espejel	18
304	Geometría Analítica I	Román Ismael Contreras Morales	19
305	Geometría Analítica I	Víctor Alberto Cruz Barriente	20
306	Geometría Analítica I	Fidencio Galicia Rodríguez	21
307	Geometría Analítica II	Pablo Rosell González	7
308	Geometría Analítica II	Álvaro Reyes García	7
309	Geometría Analítica II	Rolando Gómez Macedo	8
310	Geometría Analítica II	Juan Flores Torres	8
311	Geometría Analítica II	Jaqueline Rafaela Dolores Cañetas Ortega	8
312	Geometría Analítica II	Gabriel Gutiérrez García	10

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
313	Geometría Analítica II	Carlos Hernández Garciadiego	10
314	Geometría Analítica II	Sergio Palafox Delgado	10
315	Geometría Analítica II	Elsa Puente Vázquez	10
316	Geometría Analítica II	José Guerrero Grajeda	10
317	Geometría Analítica II	Alejandra García García	10
318	Geometría Analítica II	Francisco Manuel Barrios Panagua	10
319	Geometría Analítica II	Eugenio Garnica Vigil	12
320	Geometría Analítica II	Fernando Brambila Paz	13
321	Geometría Analítica II	Pablo Barrera Sánchez	13
322	Geometría Analítica II	Nora Isabel Pérez Quezadas	14
323	Geometría Analítica II	Tonatiuh Valdez Hernández	15
324	Geometría Analítica II	Herminio Suárez Quiroz	16
325	Geometría Analítica II	María de Lourdes Hernández Campos	17
326	Geometría Analítica II	Maya Lol Sosa Salas	17
327	Geometría Analítica II	Cristobal Falconi Hoyos	17
328	Geometría Analítica II	Camilo Camhaji García	19
329	Geometría Analítica II	Juan Manuel Martínez Nuño	20
330	Geometría Diferencial I	Pierre Michel Bayard	7
331	Geometría Diferencial I	Federico Sánchez Bringas	10
332	Geometría Diferencial I	María de los Ángeles Sandoval Romero	10
333	Geometría Diferencial I	Elhoim Llorente Sumano y Ramírez	18
334	Geometría Moderna I	Narda Cordero Michel	8
335	Geometría Moderna I	Lilia Montserrat Vite Escobedo	8
336	Geometría Moderna I	Verónica Martínez de la Vega y Mansilla	10
337	Geometría Moderna I	Francisco de Jesús Struck Chávez	10
338	Geometría Moderna I	Vinicio Antonio Gómez Gutiérrez	10
339	Geometría Moderna I	Leobardo Fernández Román	14
340	Geometría Moderna I	Natalia Jonard Pérez	17
341	Geometría Moderna I	Héctor Eduardo Martínez Moreno	18
342	Geometría Moderna I	Saúl Arce Rocha	19
343	Geometría Moderna I	Jorge Alonso Santos Mellado	19
344	Geometría Moderna II	Isabel Alicia Hubard Escalera	10
345	Geometría Moderna II	Luis Manuel Venegas Grajales	18

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
346	Inferencia Estadística	Jimmy Hernández Morales	7
347	Inferencia Estadística	Edna Gabriela López Estrada	10
348	Inferencia Estadística	Carlos Erwin Rodríguez Hernández-Vela	10
349	Inferencia Estadística	Alan Riva Palacio Cohen	12
350	Inferencia Estadística	Miguel Arturo Ballesteros Montero	13
351	Inferencia Estadística	Yolanda Martínez Guerrero	16
352	Inferencia Estadística	Antonio Soriano Flores	18
353	Inferencia Estadística	Martha Angélica Montes Fonseca	18
354	Inglés I	María Azucena Rivera Vidal	14
355	Inglés I	Paulina Torres Carranza	17
356	Inglés II	Lidia Fabiola Quevedo Rojas	8
357	Inglés II	Julio César Escorza Isguerra	11
358	Inglés II	Manuel Enrique Camargo Coronel	13
359	Inglés III	Martha del Carmen Riveroll Alvarez	7
360	Inglés III	Cesar Martinez Salazar	14
361	Inglés V	María Teresa Pozos Delgado	7
362	Inglés V	Silvia Loera Rivera	11
363	Inglés VI	Alejandro Pérez Meléndez	10
364	Inglés VI	Benito Peter Nava Kopp	14
365	Introducción a Ciencias de la Computación	Amparo López Gaona	8
366	Introducción a Ciencias de la Computación	Pedro Ulises Cervantes González	10
367	Introducción a Ciencias de la Computación	Canek Peláez Valdés	13
368	Introducción a Ciencias de la Computación	Francisco Valdés Souto	15
369	Introducción a la Física Cuántica	Sergio Alfonso Pelayo Escalera	8
370	Introducción a la Física Cuántica	Raúl Wayne Gómez González	10
371	Introducción a la Física Cuántica	Héctor Jesús Díaz Jiménez	10
372	Introducción a la Física Cuántica	Augusto Cabrera Manuel	10
373	Introducción a la Física Cuántica	Saúl Noé Ramos Sánchez	12
374	Introducción a las Matemáticas Discretas	Karina García Buendía	7
375	Introducción a las Matemáticas Discretas	Carlos Alberto Serrato Hernández	7

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
376	Introducción a las Matemáticas Discretas	José Luis Miranda Olvera	8
377	Introducción a las Matemáticas Discretas	Rita Esther Zuazua Vega	8
378	Introducción a las Matemáticas Discretas	María del Rocío Sánchez López	12
379	Introducción a las Matemáticas Discretas	Gerardo Miguel Tecpa Galván	13
380	Introducción a las Matemáticas Discretas	Mucuy-Kak del Carmen Guevara Aguirre	15
381	Introducción a las Matemáticas Discretas	Jesús Alva Samos	17
382	Investigación de Operaciones	Leonardo López Monroy	7
383	Investigación de Operaciones	Jesús Agustín Cano Garcés	8
384	Investigación de Operaciones	Nilse Pamela Romero Basurto	8
385	Investigación de Operaciones	Ana Lilia Anaya Muñoz	10
386	Investigación de Operaciones	María del Carmen Fernández García	17
387	Investigación de Operaciones	Edgar Gil Hernández Díaz	19
388	Lógica Matemática I	Luis Jesús Turcio Cuevas	8
389	Lógica Matemática I	Carlos Torres Alcaraz	13
390	Lógica Matemática I	María de la Asunción Preisser Rodríguez	17
391	Lógica Matemática II	José Gabriel Ocampo Márquez	10
392	Lógica Matemática II	María de Lourdes Guerrero Zarco	16
393	Manejo de Datos	Miguel Murguía Romero	8
394	Manejo de Datos	Jessica Santizo Galicia	8
395	Manejo de Datos	Araceli Eugenia Mercado Fernández	8
396	Manejo de Datos	Miguel Ehécatl Morales Trujillo	8
397	Manejo de Datos	Mario González Ruiz	9
398	Manejo de Datos	Manuel Alcántara Juárez	11
399	Manejo de Datos	Gildardo Bautista García Cano	12
400	Manejo de Datos	Gerardo Avilés Rosas	12
401	Manejo de Datos	José Alfredo Cobián Campos	13
402	Manejo de Datos	Virginia Teodosio Procopio	14
403	Manejo de Datos	Javier García García	16
404	Manejo de Datos	Miguel Ángel Pérez León	18
405	Manejo de Datos	Oscar Ruiz Salinas	20

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
406	Manejo de Datos	Víctor Manuel Carreón Calderón	20
407	Manejo de Datos	Benjamín Figueroa Solano	20
408	Matemáticas Actuariales del Seguro de Personas I	Elisa Sugey Hernández Castañeda	7
409	Matemáticas Actuariales del Seguro de Personas I	María Fernanda López Hernández	7
410	Matemáticas Actuariales del Seguro de Personas I	Yurguen Hugo Camargo Serafin	19
411	Matemáticas Actuariales del Seguro de Personas II	Jorge Luis Reyes García	7
412	Matemáticas Actuariales del Seguro de Personas II	Maria Isabel Servin Arellano	8
413	Matemáticas Actuariales para Seguro de Daños, Fianzas y Reaseguro	Janeth Fernández Ramos	7
414	Matemáticas Actuariales para Seguro de Daños, Fianzas y Reaseguro	Jorge Luis Reyes García	8
415	Matemáticas Actuariales para Seguro de Daños, Fianzas y Reaseguro	Elisa Sugey Hernández Castañeda	8
416	Matemáticas Actuariales para Seguro de Daños, Fianzas y Reaseguro	Irma Rocío Villa Valles	10
417	Matemáticas Actuariales para Seguro de Daños, Fianzas y Reaseguro	Juan Carlos Vargas Aguilar	14
418	Matemáticas Financieras	Irma Rocío Zavala Sierra	7
419	Matemáticas Financieras	Ana Laura Duarte Carmona	21
420	Matemáticas para las Ciencias Aplicadas I	José Luis Gutiérrez Sánchez	8
421	Matemáticas para las Ciencias Aplicadas I	Ricardo Zavaleta Madrid	8
422	Matemáticas para las Ciencias Aplicadas I	Manuel Velasco Juan	10
423	Matemáticas para las Ciencias Aplicadas I	Luis Enrique Quintanar Cortés	10
424	Matemáticas para las Ciencias Aplicadas I	Alejandro Villarreal López	12
425	Matemáticas para las Ciencias Aplicadas I	Raúl Jaime Torres Rojas	16
426	Matemáticas para las Ciencias Aplicadas II	Omar Recillas Ayala	8
427	Matemáticas para las Ciencias Aplicadas II	Arturo Velasco Pelayo	10
428	Matemáticas para las Ciencias Aplicadas II	Ricardo Zavaleta Madrid	10

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
429	Matemáticas para las Ciencias Aplicadas II	Víctor Hugo Alvarado Beceril	10
430	Matemáticas para las Ciencias Aplicadas II	Juan Carlos Balleza García	14
431	Matemáticas para las Ciencias Aplicadas II	Pedro Eduardo Miramontes Vidal	14
432	Matemáticas para las Ciencias Aplicadas II	Juan Carlos Díaz Patiño	15
433	Matemáticas para las Ciencias Aplicadas II	Héctor Jesús Díaz Jiménez	18
434	Matemáticas para las Ciencias Aplicadas II	Juan Guillermo Munguía Fernández	18
435	Matemáticas para las Ciencias Aplicadas III	Jorge Clouthier López	8
436	Matemáticas para las Ciencias Aplicadas III	Luis Enrique Quintanar Cortés	8
437	Matemáticas para las Ciencias Aplicadas III	Daniel Castillo Rodríguez Arana	8
438	Matemáticas para las Ciencias Aplicadas III	Pedro Porras Flores	10
439	Matemáticas para las Ciencias Aplicadas III	José Luis Gutiérrez Sánchez	10
440	Matemáticas para las Ciencias Aplicadas III	Jorge Avella Martínez	17
441	Matemáticas para las Ciencias Aplicadas III	Benjamín Millán Ramos	17
442	Matemáticas para las Ciencias Aplicadas IV	Mariana Soledad Centeno Sierra	10
443	Matemáticas para las Ciencias Aplicadas IV	Isaac Arelio Ríos	10
444	Matemáticas para las Ciencias Aplicadas IV	Oscar García Zarco	10
445	Matemáticas para las Ciencias Aplicadas IV	Faustino Sánchez Garduño	12
446	Matemáticas para las Ciencias Aplicadas IV	Zeus Alberto Valtierra Quintal	16
447	Matemáticas para las Ciencias Aplicadas IV	Jorge Avella Martínez	16
448	Matemáticas para las Ciencias Aplicadas IV	Jesús Edel Cereceres Delgado	16
449	Matemáticas para las Ciencias Aplicadas IV	Víctor Hugo Alvarado Beceril	19
450	Mecánica Analítica	María Luisa Marquina Fábriga	10

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
451	Mecánica Analítica	Vivianne Marquina Fábrega	10
452	Mecánica Analítica	Rosa María Méndez Vargas	10
453	Mecánica Analítica	Graciela Velasco Herrera	17
454	Mecánica Cuántica	Luis Fernando Magaña Solís	8
455	Mecánica Cuántica	Fernando Matías Moreno Yntriago	10
456	Mecánica Cuántica	Héctor Hernández Coronado	10
457	Mecánica Cuántica	Enriqueta Hernández Saldaña	10
458	Mecánica Cuántica	Carlos Ramírez Ramos	15
459	Mecánica Vectorial	Isaías Rodríguez Aguirre	7
460	Mecánica Vectorial	Roberto Alejandro Ruelas Mayorga	8
461	Mecánica Vectorial	Eric Vázquez Jauregui	8
462	Mecánica Vectorial	Manuel Gerardo Quintana García	8
463	Mecánica Vectorial	José Antonio Fragoso Uroza	8
464	Mecánica Vectorial	Vicenta Sánchez Morales	10
465	Mecánica Vectorial	Jerónimo Alonso Cortez Quezada	10
466	Mecánica Vectorial	Juan Manuel Eugenio Ramírez de Arellano Niño-Rincón	10
467	Mecánica Vectorial	Darío Núñez Zúñiga	10
468	Mecánica Vectorial	Ramón Peralta y Fabi	10
469	Mecánica Vectorial	Susana Orozco Segovia	10
470	Mecánica Vectorial	Luciano Martínez Balbuena	10
471	Mecánica Vectorial	Jorge Ramón Soto Mercado	10
472	Mecánica Vectorial	Víctor Hugo Meza Laguna	10
473	Mecánica Vectorial	Consuelo García Alcántara	11
474	Mecánica Vectorial	José Alfredo Heras Gómez	13
475	Mecánica Vectorial	Roberto de Jesús León Montiel	14
476	Mecánica Vectorial	Francisco Javier Mandujano Sánchez	16
477	Mecánica Vectorial	Francisco Espinosa Magaña	16
478	Mecánica Vectorial	David García Gudiño	18
479	Mercados Financieros y Valuación de Instrumentos	José David Santos Salazar	7
480	Mercados Financieros y Valuación de Instrumentos	Jesús Abraham Cantú Orozco	8
481	Mercados Financieros y Valuación de Instrumentos	Alma Rosa Bustamante García	15
482	Mercados Financieros y Valuación de Instrumentos	Angélica Vargas Serrano	19
483	Métodos Cuantitativos en Finanzas	Enrique Maturano Rodríguez	7

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
484	Métodos Cuantitativos en Finanzas	Lizeth Fabiola Sánchez Suárez	7
485	Métodos Cuantitativos en Finanzas	Nayeli Castillo Carranco	7
486	Métodos Cuantitativos en Finanzas	Jorge Luis Silva Haro	8
487	Métodos Cuantitativos en Finanzas	Alberto Cadena Martínez	16
488	Métodos Cuantitativos en Finanzas	Daniela Alejandra González Ramírez	18
489	Métodos Cuantitativos en Finanzas	Juan Diego Nieves Ledesma	19
490	Modelado y Programación	Ana Libia Eslava Cervantes	8
491	Modelado y Programación	Odín Miguel Escorza Soria	10
492	Modelado y Programación	José de Jesús Galaviz Casas	13
493	Modelos de Supervivencia y de Series de Tiempo	Gerardo Sisniega Lira	7
494	Modelos de Supervivencia y de Series de Tiempo	Ángel Manuel Godoy Aguilar	7
495	Modelos de Supervivencia y de Series de Tiempo	David Chaffrey Moreno Fernández	8
496	Modelos de Supervivencia y de Series de Tiempo	Ricardo Ramírez Aldana	10
497	Modelos de Supervivencia y de Series de Tiempo	Claudia González González	10
498	Modelos de Supervivencia y de Series de Tiempo	Margarita Elvira Chávez Cano	10
499	Modelos de Supervivencia y de Series de Tiempo	Luis Antonio Rincón Solís	10
500	Modelos de Supervivencia y de Series de Tiempo	Alberto Contreras Cristán	12
501	Modelos de Supervivencia y de Series de Tiempo	Guillermina Eslava Gómez	13
502	Modelos de Supervivencia y de Series de Tiempo	Rosa Daniela Chávez Aguilar	18
503	Modelos de Supervivencia y de Series de Tiempo	Claudia Lara Pérez Soto	18
504	Modelos de Supervivencia y de Series de Tiempo	Francisco Sánchez Villarreal	18
505	Modelos no Paramétricos y de Regresión	Jaime Vázquez Alamilla	8
506	Modelos no Paramétricos y de Regresión	Margarita Elvira Chávez Cano	9
507	Modelos no Paramétricos y de Regresión	Inocencio Rafael Madrid Ríos	9
508	Modelos no Paramétricos y de Regresión	Jose Salvador Zamora Muñoz	10

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
509	Modelos no Paramétricos y de Regresión	Javier Santibañez Cortés	10
510	Modelos no Paramétricos y de Regresión	Lizbeth Naranjo Albarrán	11
511	Modelos no Paramétricos y de Regresión	Alejandro Martínez López	19
512	Modelos no Paramétricos y de Regresión	Claudia Lara Pérez Soto	19
513	Modelos no Paramétricos y de Regresión	Rosa Daniela Chávez Aguilar	20
514	Muestreo	José Antonio Flores Díaz	11
515	Óptica	José Ignacio Jiménez Mier y Terán	8
516	Óptica	José Luis Pérez Mazariego	10
517	Óptica	Nadia Ramírez Cruz	10
518	Óptica	Adriano Valdés Gómez	12
519	Óptica	Víctor Hugo Meza Laguna	14
520	Óptica	Ali Michel Angulo Martínez	16
521	Pensiones Privadas	José Antonio Valencia Trujillo	7
522	Pensiones Privadas	César Adrián Hernández Vi-veros	7
523	Pensiones Privadas	Jonathan González Salgado	8
524	Pensiones Privadas	Ernesto Rosas Garcia	18
525	Pensiones Privadas	Francisco Fernando Morales Castro	18
526	Planeación Estratégica	Mónica Iliana Sánchez Zara-goza	7
527	Planeación Estratégica	Felipe Zamora Ramos	8
528	Probabilidad I	Jesús Alberto Rodríguez Sánchez	7
529	Probabilidad I	Laura Nayeli Liljehult León	8
530	Probabilidad I	Guadalupe Carrasco Licea	8
531	Probabilidad I	Francisco Pérez Carbajal	10
532	Probabilidad I	Perla Marysol Ruiz Arias	10
533	Probabilidad I	Yuri Salazar Flores	10
534	Probabilidad I	María Asunción Begoña Fernández Fernández	10
535	Probabilidad I	Bibiana Obregón Quintana	12
536	Probabilidad I	Jaime Vázquez Alamilla	12
537	Probabilidad I	Julio César Galindo López	16
538	Probabilidad I	Hugo Villaseñor Hernández	17
539	Probabilidad II	Daniel Alejandro Zurita Gu-tiérrez	8

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
540	Probabilidad II	Daniel Cervantes Filoteo	8
541	Probabilidad II	Jorge Luis Apátiga Sánchez	9
542	Probabilidad II	Alberto Contreras Cristán	10
543	Probabilidad II	David Josafat Santana Cobian	10
544	Probabilidad II	Sandra Palau Calderón	10
545	Probabilidad II	Fernando Guerrero Poblete	10
546	Probabilidad II	Marco Arieli Herrera Valdez	10
547	Probabilidad II	Ricardo Hoyos Argüelles	11
548	Probabilidad II	Luis Antonio Rincón Solís	11
549	Probabilidad II	Francisco Daniel Ramírez Calixto	13
550	Probabilidad II	Iván Ixcoatl Juárez López	14
551	Probabilidad II	Sergio Iván López Ortega	15
552	Probabilidad II	Miguel Ángel García Álvarez	17
553	Probabilidad II	Juan González Hernández	18
554	Probabilidad II	Fernando Díaz López	19
555	Procesos Estocásticos I	Sandra Palau Calderón	8
556	Procesos Estocásticos I	Adrián González Casanova Soberon	8
557	Procesos Estocásticos I	María Clara Fittipaldi	10
558	Procesos Estocásticos I	Jorge Ignacio González Cázares	10
559	Procesos Estocásticos I	Rodrigo Quijón Hipólito	10
560	Procesos Estocásticos I	Ricardo Hoyos Argüelles	10
561	Procesos Estocásticos I	Arrigo Coen Coria	10
562	Procesos Estocásticos I	Abraham Ramírez Hernández	18
563	Procesos Estocásticos I	Francisco Daniel Ramírez Calixto	19
564	Procesos Estocásticos I	Sergio Iván López Ortega	19
565	Productos Financieros Derivados	Alejandro Hugo Velázquez Roldán	7
566	Productos Financieros Derivados	Guillermo Rolando Rosas Figueroa	8
567	Productos Financieros Derivados	Jesús Agustín Cano Garcés	10
568	Productos Financieros Derivados	Germán Valle Trujillo	19
569	Productos Financieros Derivados	Christian Gabriel Miranda Ruíz	20
570	Programación	Miguel Murguía Romero	7
571	Programación	Armando Orozco Cortés	7
572	Programación	Benjamín Figueroa Solano	8
573	Programación	María Fernanda Sánchez Puig	8
574	Programación	Alejandro Carrillo Nolazco	8
575	Programación	Sonia Josefina Valery Lobo	9

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
576	Programación	Luis Enrique Serrano Gutiérrez	10
577	Programación	Jaime Ayala Pérez	11
578	Programación	Antonio Carrillo Ledesma	12
579	Programación	Roberto Méndez Méndez	12
580	Programación	Amparo López Gaona	14
581	Programación	Pablo Alberto Tintor Jiménez	18
582	Programación	Mario González Ruiz	20
583	Relatividad	Saúl Noé Ramos Sánchez	8
584	Relatividad	Enrique Moreno Méndez	8
585	Relatividad	Juan Carlos Degollado Daza	8
586	Relatividad	César Simón López Monsalvo	10
587	Relatividad	Juan Manuel García Islas	10
588	Relatividad	Mariano Chernicoff Minsberg	10
589	Relatividad	Yuri Bonder Grimberg	12
590	Relatividad	Roberto Allan Sussman Livovovsky	12
591	Relatividad	Eduardo Peinado Rodríguez	14
592	Relatividad	Gustavo Alfredo Arciniega Durán	18
593	Relatividad	Pedro Antonio Sánchez Serrano	18
594	Seguridad Social	Luis Alberto Galván Contreras	7
595	Seguridad Social	Daniela Alejandra González Ramírez	8
596	Seguridad Social	María del Pilar Alonso Reyes	10
597	Seminario de Álgebra A	Juan Orendain Almada	8
598	Seminario de Álgebra A	Emilio Esteban Lluis Puebla	10
599	Seminario de Álgebra A	Luis Eduardo García Hernández	15
600	Seminario de Álgebra A	Javier Valdez Quijada	20
601	Seminario de Aplicaciones Actuariales	Eduardo Torres Luna	7
602	Seminario de Aplicaciones Actuariales	Mónica Iliana Sánchez Zágoza	8
603	Seminario de Aplicaciones Actuariales	Víctor Manuel Solís Nájera	8
604	Seminario de Aplicaciones Actuariales	Alonso Baranda Lozada	8
605	Seminario de Aplicaciones Actuariales	Carlos Díaz Ávalos	10
606	Seminario de Aplicaciones Actuariales	Ricardo Ramírez Aldana	11
607	Seminario de Aplicaciones Actuariales	Israel Emmanuel Ambriz Lobato	19
608	Seminario de Aplicaciones Actuariales	Omar de la Riva Torres	20

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
609	Seminario de Aplicaciones Actuariales	José Fernando Soriano Flores	21
610	Seminario de Apoyo a la Titulación en Matemáticas	Natalia Bárbara Mantilla Beniers	15
611	Seminario de Apoyo a la Titulación en Matemáticas	Francisco Javier Torres Ayala	15
612	Seminario de Ciencias de la Computación	Jonathan Banfi Vázquez	8
613	Seminario de Ciencias de la Computación	German Ernesto Zapata Ledesma	8
614	Seminario de Ciencias de la Computación	Adriana Ramírez Vigueras	9
615	Seminario de Ciencias de la Computación	Armando Castañeda Rojano	10
616	Seminario de Ciencias de la Computación	Jorge Luis Ortega Arjona	10
617	Seminario de Ciencias de la Computación	Salvador López Mendoza	10
618	Seminario de Ciencias de la Computación	José David Flores Peñaloza	10
619	Seminario de Ciencias de la Computación	Manuel Cristobal López Michelone	14
620	Seminario de Ciencias de la Computación	Luis Alberto Ramírez Bermúdez	17
621	Seminario de Ciencias de la Computación	Josafat Isai Guerrero Iñiguez	18
622	Seminario de Estadística I	Rafael Miranda Cordero	10
623	Seminario de Estadística I	Ruth Selene Fuentes García	11
624	Seminario de Geometría A	Pablo Rosell González	8
625	Seminario de Geometría A	Antonio Lascurain Orive	14
626	Seminario de Geometría A	José Collins Castro	17
627	Seminario de Topología A	Alejandro Illanes Mejía	10
628	Seminario de Topología A	Ernesto Alejandro Vázquez Navarro	14
629	Seminario de Topología A	Enrique Guillermo Bazúa Durán	19
630	Seminario Matemáticas Aplicadas I	Hortensia Galeana Sánchez	10
631	Seminario Matemáticas Aplicadas I	María de la Luz Jimena de Teresa de Oteyza	10
632	Seminario Matemáticas Aplicadas I	Adrián González Casanova Soberon	10
633	Seminario Matemáticas Aplicadas I	Karla Genoveva Bassols Belllo	14
634	Seminario Matemáticas Aplicadas I	Pedro José Sobrevilla Moreno	18

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
635	Seminario Matemáticas Aplicadas I	José Guadalupe Vázquez Vázquez	20
636	Seminario sobre Enseñanza de las Matemáticas I	Carmen Martínez-Adame Isaís	10
637	Seminario sobre Enseñanza de las Matemáticas I	José Rafael Martínez Enríquez	10
638	Seminario sobre Enseñanza de las Matemáticas I	Adriana Léon Montes	14
639	Seminario sobre Enseñanza de las Matemáticas I	Antonio García Flores	15
640	Seminario sobre Enseñanza de las Matemáticas I	Vicente Carrión Velázquez	19
641	Taller de Herramientas Computacionales	Armando Orozco Cortés	8
642	Temas Selectos de Economía	Luis Alberto Quezada Téllez	8
643	Teoría de Gráficas	María del Pilar Valencia Saravia	7
644	Teoría de la Medida I	María Emilia Caballero Acosta	10
645	Teoría de los Conjuntos I	David Meza Alcántara	8
646	Teoría de los Conjuntos I	Ángel Tamariz Mascarúa	13
647	Teoría de los Conjuntos I	Rodrigo Edmundo Cepeda Morales	15
648	Teoría de los Conjuntos I	Erick García Ramírez	19
649	Teoría de los Números I	Alejandro De Las Peñas Castaño	8
650	Teoría de los Números I	Elhoim Llorente Sumanoy Ramírez	20
651	Teoría de los Números II	Ernesto Cruz Guerrero	8
652	Teoría de Redes	María Isabel Escalante Membrillo	17
653	Teoría del Riesgo	Luis Alberto Olvera García	7
654	Teoría del Riesgo	Alejandro Santoyo Cano	8
655	Teoría del Riesgo	Roberto Méndez Rosas	17
656	Teoría del Riesgo	Yuri Salazar Flores	19
657	Teoría del Seguro	María Patricia Luna Díaz	8
658	Teoría del Seguro	Alfonso Parrao Guzman	19
659	Termodinámica	René Carrillo Moreno	8
660	Termodinámica	María Michel Duque Vega	8
661	Termodinámica	Rodolfo Patricio Martínez y Romero	10
662	Termodinámica	Víctor Manuel Romero Rochín	10

La tabla continúa en la siguiente página

	Materia	Profesor	Horario
663	Termodinámica	Manuel Gerardo Quintana García	10
664	Termodinámica	Ignacio Campos Flores	16
665	Termodinámica	Carlos Enrique Román Velázquez	20
666	Topología I	Gerardo Acosta García	8
667	Topología I	Sergio Macías Álvarez	10
668	Topología I	Ángel Tamariz Mascarúa	10
669	Topología I	Manuel Alejandro Lara Mary	12
670	Topología I	María Isabel Puga Espinosa	13
671	Topología I	Patricia Pellicer Covarrubias	16
672	Topología I	Jorge Moreno Montes	19
673	Topología III	Alejandro Illanes Mejía	9
674	Variable Compleja I	Emigdio Martínez Ojeda	10
675	Variable Compleja I	Héctor Fidencio Sánchez Morgado	10
676	Variable Compleja I	Oscar Alfredo Palmas Velasco	10
677	Variable Compleja I	Roberto Pichardo Mendoza	12
678	Variable Compleja I	Carisa Cano Figueroa	13
679	Variable Compleja I	Micho Durdevich Lucic	16
680	Variable Compleja I	Mico Djurdjevic	16
681	Variable Compleja I	Alejandro Gaona Ordoñez	18
682	Variable Compleja I	Jesús Manuel Mayorquín García	19

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 *checha_ind_materia* se encarga de obtener los 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 excepciones 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 inglés 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” utilizando 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.

The screenshot shows a webpage titled "Horarios 2021-1" from the Faculty of Sciences. The page includes navigation links for "Cienencias", "Horarios 2021-1", "Actuaría", and "Modelos no Paramétricos y de Regresión". Below the title, it says "Actuaría (plan 2015)". A search bar is present with the placeholder "Buscar en grupos". The main content lists several groups with their respective professors and schedules:

- Grupo 9220**: Profesor Jaime Vázquez Alamilla (lu mi vi 10 a 11), Ayudante Omar Rodríguez Torres (ma ju 10 a 11), Presentación.
- Grupo 9222**: Profesor María del Pilar Alonso Reyes (lu mi vi 12 a 13), Ayudante Luisa Téllez Moreno (ma ju 12 a 13), Presentación.
- Grupo 9224**: Profesor Gonzalo Pérez de la Cruz (lu mi vi 17 a 18), Ayudante Dioney Alonso Rosas Sánchez (ma ju 17 a 18).
- Grupo 9226**: Profesor Francisco Sánchez Villarreal (lu mi vi 18 a 19), Ayudante Santiago Lara Jiménez (ma ju 18 a 19), Presentación.

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

23. Notas de T26

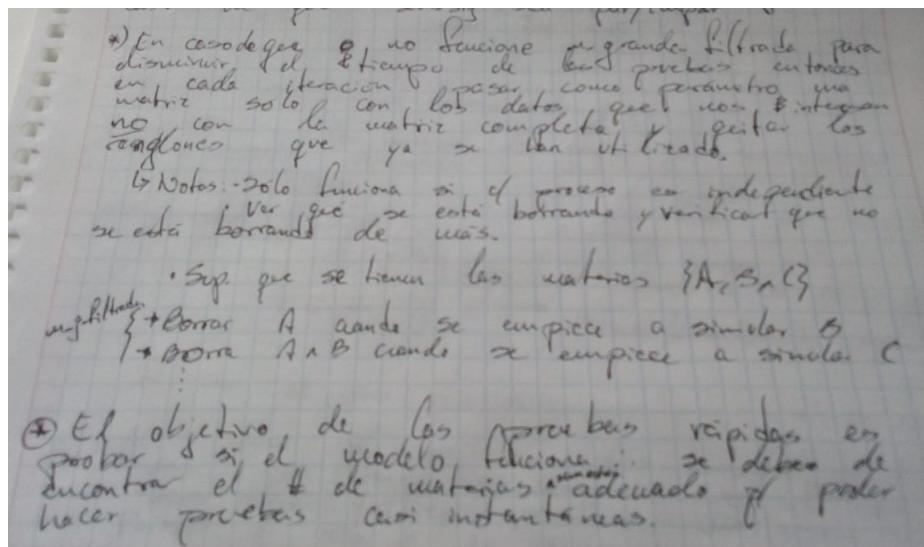


Figura E.3: Notas de T26

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 generar 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 empalan clases)
37. Para la elección de q_1 y q_2 se debe darle prioridad a la varianza no al mínimo y al máximo porque se pueden tener casos en los que el mínimo y el máximo estén muy cercanos a cero (gráfica superior) pero su varianza es grande. Queremos que la varianza se encuentre alrededor del cero (gráfica inferior).

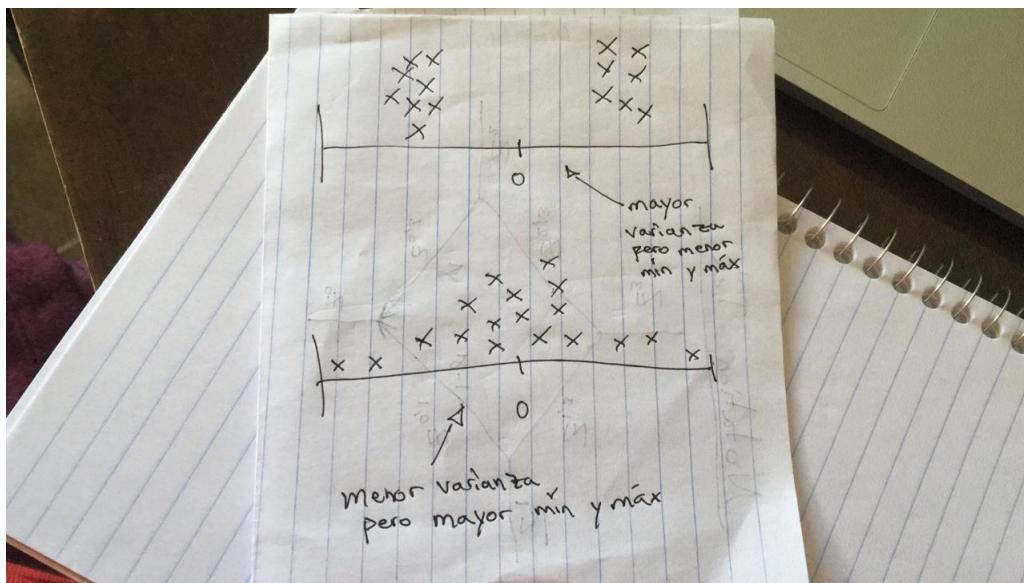


Figura E.5: Ejemplo de varianza

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>

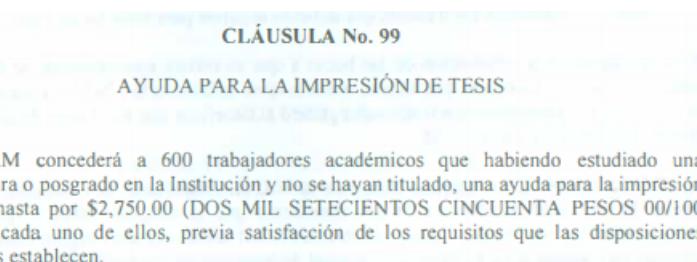


Figura E.6: *Cláusula 99 CCTPA: Ayuda para la impresión de la tesis*

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_{32xx})
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 `\subsubsectionname{\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>.

— Planes de Estudio —

Llene los datos que se le piden a continuación:

Consulta de Planes de Estudios	
CLAVE DEL PLAN DE ESTUDIO	
Estructura	Consultar

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 $\text{sqrt}(|\text{residuals}|)$ 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:

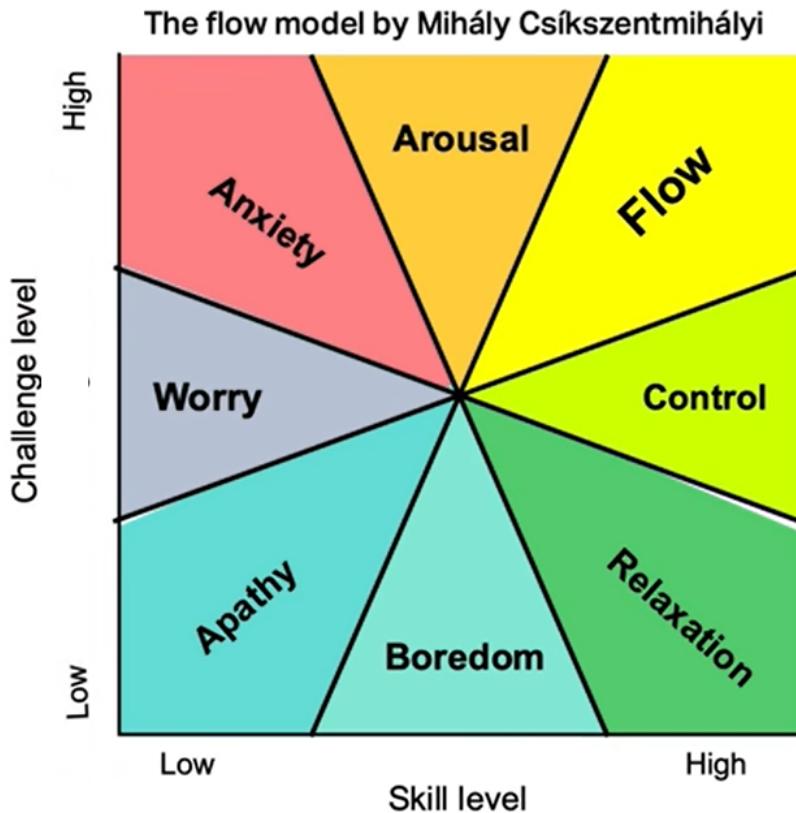
- RAE: <https://www.rae.es/dpd/may%C3%BAsculas#33c>
- Otro: <http://iesbinef.educa.aragon.es/lengua/ortografia/reglas/reglama.htm>

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. El comando `ctrl + shift + f`, de *R*, busca en todos los archivos de una carpeta.
- 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*, *Econometrica*, 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