



# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

---

## FACULTAD DE CIENCIAS

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

## T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIA

PRESENTA:

MIRIAM GABRIELA COLÍN NÚÑEZ

TUTOR

DR. ARRIGO COEN CORIA

CIUDAD UNIVERSITARIA, CD. MX., 2020





**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:****Datos del sinodal 2:****Datos del sinodal 3:****Datos del sinodal 4:****Datos del sinodal 5:****Datos del trabajo escrito:**

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

(Número de Páginas)

2020

*Dedicado a*

# Agradecimientos

¡Muchas gracias a todos!

# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Motivación . . . . .	1
1.2. Definición de conceptos . . . . .	2
1.3. Nomenclatura . . . . .	2
1.4. Planteamiento del problema . . . . .	3
1.5. Objetivos . . . . .	5
1.6. Datos a analizar . . . . .	5
<b>2. Extracción de datos</b>	<b>9</b>
2.1. Extracción de información con la aplicación SelectorGadget . . . . .	10
2.2. Estructura de las páginas web . . . . .	10
2.3. Tipos de grupos de las páginas web de la Facultad de Ciencias . . . . .	12
2.4. Limpieza de base de datos . . . . .	13
2.4.1. Problemas de falta de información . . . . .	13
2.4.2. Problemas de información repetida . . . . .	15
2.4.3. Otros problemas al extraer información . . . . .	17
2.5. Matrices de datos . . . . .	19
<b>3. Análisis estadístico</b>	<b>23</b>
3.1. Análisis estadístico básico . . . . .	24
3.2. Análisis estadístico por grupo de datos . . . . .	28
3.3. Distribución del tamaño de los grupos . . . . .	32
3.4. Comportamientos por hora . . . . .	34
<b>4. Simulación</b>	<b>37</b>
4.1. Funciones hechas en R . . . . .	37
4.2. Obtención de los parámetros $q_1$ y $q_2$ . . . . .	38
4.3. Simulación de la demanda de alumnos . . . . .	43
4.4. Simulación de tamaño de grupos . . . . .	44
4.5. Obtención de información para solicitudes . . . . .	44
4.6. Simulación de solicitudes de profesores oculta y pseudo-real . . . . .	47
4.6.1. Simulación de elección de materia . . . . .	47
4.6.2. Simulación de elección de horario . . . . .	47
4.7. Simulación de esqueletos . . . . .	47
4.8. Calificación de esqueletos de horario . . . . .	48
4.9. Generación de asignaciones . . . . .	49

4.9.1. Calificación de asignaciones de grupo . . . . .	49
<b>5. Teoría del Algoritmo Genético aplicado a los horarios</b>	<b>51</b>
5.1. Ciclo de la evolución natural . . . . .	52
5.1.1. Selección . . . . .	53
5.1.2. Cruce . . . . .	53
5.1.3. Mutación . . . . .	53
5.1.4. Reemplazamiento . . . . .	53
5.2. Algoritmo Genético aplicado a la generación de esqueletos de horario . . . .	53
5.3. Algoritmo Genético aplicado a la generación de asignaciones de grupos . . .	53
<b>6. Resultados del Algoritmo Genético</b>	<b>55</b>
<b>7. Comportamiento de la selección</b>	<b>57</b>
<b>8. Conclusiones</b>	<b>59</b>
<b>Apéndices</b>	<b>63</b>
8.1. Observaciones / Notas . . . . .	63
8.2. Resultados útiles . . . . .	69
8.3. Abreviaturas . . . . .	70

# Índice de figuras

1.1. Número de alumnos totales por semestres pares e impares . . . . .	6
1.2. Histograma del número de alumnos por semestre . . . . .	7
1.3. Número de alumnos por turno . . . . .	8
1.4. Histograma del número de alumnos por turno . . . . .	8
2.1. Página de horarios de la FC . . . . .	9
2.2. Aplicación SelectorGadget . . . . .	10
2.3. Tipo de grupo A . . . . .	12
2.4. Tipo de grupo B . . . . .	13
2.5. Tipo de grupo C . . . . .	13
2.6. Ejemplo de página web en blanco . . . . .	14
2.7. Ejemplo de grupo sin información de salón . . . . .	14
2.8. Ejemplo de grupo sin información de alumnos . . . . .	15
2.9. Ejemplo de grupo sólo con horario . . . . .	15
2.10. Ejemplo de información repetida: Planes de estudio . . . . .	16
2.11. Ejemplo de información repetida: Materia con nombres distintos . . . . .	16
2.12. Ejemplo de información repetida: Mismo profesor, materias distintas . . . . .	17
2.13. Ejemplo de grupo con un alumno . . . . .	17
2.14. Ejemplo de grupo con medias horas . . . . .	18
2.15. Ejemplo de grupo con horarios múltiples . . . . .	18
2.16. Ejemplo de grupo de inglés . . . . .	18
2.17. Ejemplo de grupo con estructura diferente . . . . .	19
3.1. Número total de alumnos por semestre . . . . .	25
3.2. Media de alumnos por semestre . . . . .	26
3.3. Desviación estándar del número de alumnos por semestre . . . . .	27
3.4. Descomposición por el método aditivo de Holt-Winters: Total de alumnos por semestre . . . . .	28
3.5. Número de alumnos de semestres pares e impares . . . . .	29
3.6. Histograma del número de alumnos de semestres pares e impares . . . . .	30
3.7. Número de alumnos por turno de todos los semestres . . . . .	31
3.8. Histograma del número de alumnos de los turnos matutino y vespertino . . . . .	32
3.9. Histograma del número de alumnos por grupo de todos los semestres . . . . .	33
3.10. Densidades del número de alumnos por grupo de cada semestre . . . . .	34
3.11. Número promedio de grupos por hora . . . . .	35
3.12. Número promedio de alumnos por hora . . . . .	36



4.1.	<i>Diagrama de flujo de la función <b>gen_asignacion</b></i>	38
4.2.	<i>Matriz con medidas de dispersión</i>	39
4.3.	<i>Promedio de la desviación estándar: 5 materias, 12 pruebas</i>	40
4.4.	<i>Promedio de la desviación estándar: 10 materias, 6 pruebas</i>	40
4.5.	<i>Promedio de la desviación estándar: 10 materias, 4 pruebas</i>	41
4.6.	<i>Promedio de la desviación estándar: 5 materias, 4 pruebas</i>	41
4.7.	<i>Diagrama de los intervalos de confianza</i>	42
4.8.	<i>Matriz con medidas de dispersión de prueba aleatoria</i>	42
4.9.	<i>Matriz con alumnos corregidos</i>	43
4.10.	<i>Vector con demanda simulada para el 2020-2</i>	43
4.11.	<i>Profesores de tiempo completo: SelectorGadget</i>	45
4.12.	<i>Vector de profesores de tiempo completo</i>	46
5.1.	<i>Algoritmo Genético</i>	52
5.2.	<i>Algoritmo Genético aplicado</i>	53
8.1.	<i>ITAM Probabilidad I</i>	61
8.2.	<i>Resumen de clases de ingles antes de modificación</i>	64
8.3.	<i>Ejemplo de horarios de semestre 2021-1</i>	66
8.4.	<i>Notas de T26</i>	66
8.5.	<i>Ejemplo de Roxygen</i>	67
8.6.	<i>Ejemplo de varianza</i>	68
8.7.	<i>Cláusula 99 CCTPA: Ayuda para la impresión de la tesis</i>	69

# Índice de tablas

1.1.	<i>Ejemplo de asignación</i>	5
1.2.	<i>Grupos de datos</i>	7
2.1.	<i>Planes de estudio por carrera con clave</i>	11
2.2.	<i>Descripción de columnas de mat_posibles_url</i>	12
2.3.	<i>Descripción de columnas de m_grande</i>	21
4.1.	<i>Posibles valores para <math>q_1</math> y <math>q_2</math></i>	39
4.2.	<i>Diferencias en nombres de profesores de tiempo completo</i>	46
8.1.	<i>Abreviaturas</i>	70



# Códigos

8.1. <i>Ejemplo de ciclo for</i> . . . . .	64
8.2. <i>Ejemplo de estructura de funciones</i> . . . . .	65



# 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, de los cuales se obtendrá un número estimado de alumnos, para cada materia de las carreras, del Departamento de Matemáticas, con el cual se podrán hacer aleatoriamente esqueletos de horarios que se calificarán de acuerdo a dicha demanda 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 futuros alumnos y la (posible) disminución del tiempo que toma realizar los esqueletos y la asignación de profesores en la Facultad.

Actualmente para hacer la asignación de horarios primero se reúne el comité encargado de dicha tarea para 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 elegido, 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 base y los de asignatura. Los profesores de base, por contrato, deben de cubrir ciertas horas de clase por lo que al momento de hacer la asignación, éstos tienen prioridad sobre los profesores de asignatura. 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 en caso de que el cupo del salón lo permita, 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 de Ciencias por algún profesor.

**Horario:** Hora en la que se imparte alguna materia.

**Esqueleto:** Conjunto Materia-Horario.

**Asignación:** Conjunto Materia-Horario-Profesor.

**Grupo:** Clave con la que se identifica una asignación.

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

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

## 1.3. Nomenclatura

$m$  : Número de materias que se van a impartir

$p$  : Número de profesores que van impartir alguna materia

$c$  : Número de salones de clase que se pueden ocupar

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

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

$k$  : Índice para salones,  $k \in \{1, 2, 3, \dots, c\}$

$h$  : Índice para las horas del día,  $h \in \{1, 2, 3, \dots, 15\}$

$U_{j,i,h,k}$  : Utilidad de que el profesor  $i$  imparta la materia  $j$  en el salón  $k$  a la hora  $h$

$x_{j,i,h,k}$  : Variable binaria que vale 1 si la materia  $j$  es impartida por el profesor  $i$  en el salón  $k$  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

$V_{j,k}$  : Variable binaria que vale 1 si la materia  $j$  puede impartirse en el salón  $k$  y cero en otro caso

$s$  : Semestre a simular

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

$m\_grande$  : Matriz en la que se guarda la información por semestres

$r$  : Matriz  $m\_filtrada$ , submatriz de  $m\_grande$

$vec\_sem\_sig$ : Vector con los semestres que se van a simular

$X_4$  : Analizar presentación: Hacer varias pruebas con distintas combinaciones y elegir el mejor estilo/presentación

$X_{14}$  : Revisar/Investigar al respecto del problema y resolverlo

$num\_sim$  : Número de simulaciones de la demanda de alumnos para  $s$

:

:

:

:

:

## 1.4. Planteamiento del problema

En el problema de asignación de horarios se quiere asociar un profesor con una materia, un salón y un horario. Existen trabajos que han abordado este problema desde otro punto de vista, por ejemplo Yazdani, Naeri y Zeinali, en su artículo *Algorithms for university course scheduling problems* (8), proponen un modelo en el cual se toman 2 decisiones: la asignación de profesor por materia y el salón en el cual se va a impartir cada materia.

Con la función objetivo planteada en dicho modelo se desea maximizar la utilidad de que el profesor  $i$  imparta la materia  $j$ , más la utilidad de que el profesor  $i$  dé clases el día  $t$ , más la utilidad de la materia  $j$  por ser impartida en el día  $t$ .

Dos diferencias principales entre su modelo y el que proponemos en este trabajo son:

- 1) No tomamos en cuenta el día en el que se imparte la materia porque suponemos que todas las materias se imparten de lunes a viernes, a la misma hora, en el mismo salón.
- 2) Deseamos maximizar la utilidad de que el profesor  $i$  imparta la materia  $j$  en el salón  $k$  a la hora  $h$ .

Los elementos que consideramos en nuestro modelo son:

- Esqueletos de horario: Matriz de 15 renglones con las horas (7-8, 8-9, ..., 21-22) y tantas columnas como materias. En la entrada  $(i, j)$  contiene el número de grupos simulados de la  $i$ -ésima hora 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 FC.
- Conjunto de profesores: Nombres de profesores de base y de asignatura.



I) Variables de decisión:

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

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

$$\text{máx } z = \sum_{i=1}^p \sum_{j=1}^m \sum_{k=1}^c \sum_{h=1}^{15} x_{j,i,h,k} U_{j,i,h,k} \text{ s. a}$$

III) Restricciones:

$$\sum_{i=1}^p \sum_{k=1}^c \sum_{h=1}^{15} x_{j,i,h,k} = 1 \quad \forall j \quad (1.1)$$

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

$$\sum_{i=1}^p \sum_{j=1}^m x_{j,i,h,k} \leq 1 \quad \forall k, h \quad (1.3)$$

$$\sum_{k=1}^c \sum_{h=1}^{15} x_{j,i,h,k} \leq V_{i,j} \quad \forall i, j \quad (1.4)$$

$$\sum_{i=1}^p \sum_{h=1}^{15} x_{j,i,h,k} \leq V_{j,k} \quad \forall j, k \quad (1.5)$$

$$x_{j,i,h,k}, V_{j,i}, V_{j,k} \in \{0, 1\} \quad \forall j, i, h, k \quad (1.6)$$

Con las restricciones del tipo (1) aseguramos que todas las materias sean dadas. Con las del tipo (2) se asegura que cada profesor no tenga más de un curso por hora. Las restricciones del tipo (3) nos indican que se puede impartir máximo una materia en cada salón por hora. Con las del tipo (4) aseguramos que los profesores tengan asignadas materias que puedan impartir. Las restricciones del tipo (5) especifican que cada materia es asignada a un salón donde pueda impartirse esa clase. Finalmente con las restricciones del tipo (6) se especifica que las variables utilizadas son binarias.

En el planteamiento se tienen dos tipos de restricciones: duras y suaves, las restricciones duras son las que nos permiten tener soluciones factibles al cumplirlas en su totalidad y las restricciones suaves nos permiten evaluar la calidad de las diferentes soluciones, usualmente están asociadas a preferencias y se cumplen en la medida de lo posible, pero no afectan la factibilidad de las soluciones.

El conjunto de soluciones se presenta por medio de la matriz *mat\_asignaciones* la cual es una matriz de cuatro columnas y tantos renglones como materias se tengan. En el *i*-ésimo renglón se tiene la información de la *i*-ésima materia con su respectivo profesor, horario y salón asignados. En la tabla 1.1 se muestra un ejemplo del resultado de la asignación.

Materia	Profesor	Horario	Salón
Inferencia Estadística	Margarita Chávez Cano	9-10	O123
Estadística II	Jaime Vázquez Alamilla	10-11	001 (Yelizcalli)
Estadística Bayesiana	Ruth Fuentes García	11-12	P110
Estadística III	Lizbeth Naranjo Albarrán	13-14	Aula Magna P

Tabla 1.1: Ejemplo de asignación

## 1.5. Objetivos

El primer objetivo del trabajo es hacer dos funciones, una que genere esqueletos de horarios y otra que genere una asignación de profesores por materia y horario que cubra la demanda de alumnos estimada para un semestre dado.

Los esqueletos de horarios son utilizados para simular una posible elección de materias y horarios de los profesores para finalmente hacer la asignación correspondiente a la hora, materia, profesor y salón de cada grupo.

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 un análisis adecuado de los datos, se hicieron dos divisiones de ellos. La primera fue con respecto al tipo de semestre, par o impar y la segunda se hizo con respecto al turno, matutino o vespertino.

Se tomó la información de la materia de *Probabilidad I*, desde el semestre 2015-1 hasta el 2020-1, para explicar la elección de las divisiones. Cabe aclarar que dicha materia en la carrera de Actuaría es una materia obligatoria de tercer semestre.

Veamos primero la división de acuerdo a semestres pares e impares.

En la figura 1.1 se tiene una gráfica en la 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. Se observa que en todo momento el número de alumnos de los semestres impares es mayor al número de alumnos de los semestres pares. Lo cual nos interesa porque al momento de simular se debe de tomar en cuenta que siempre el número de alumnos totales de semestres impares debe de ser mayor al número total de alumnos de los semestres pares.

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

Las líneas que se encuentran sobre los histogramas son densidades estimadas que se ajustan a los histogramas. Algunos datos que se pueden obtener de esas densidades 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.

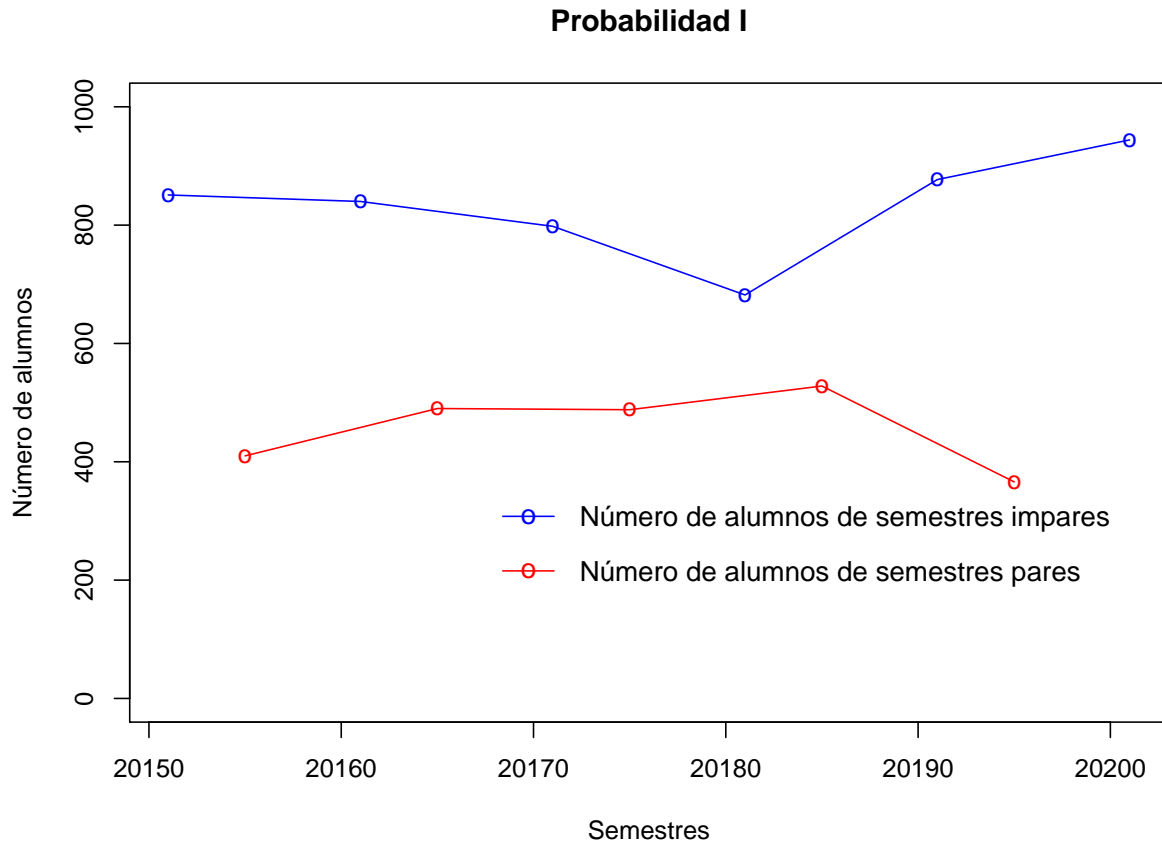


Figura 1.1: *Número de alumnos totales por semestres pares e impares*

Ahora veamos la segunda división de acuerdo a los turnos matutino y vespertino.

En la gráfica de 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, esto impacta en el hecho de que por semestres la varianza en el turno matutino es mucho mayor que en el turno vespertino lo cual indica que en el turno vespertino se tiene prácticamente el mismo número de alumnos sin importar si la materia pertenece a un semestre par o impar, a diferencia de lo que ocurre en el turno matutino en donde si influye el hecho de que la materia corresponda a un semestre par o impar.

Al igual que en la primera división se obtuvo una gráfica que contiene dos histogramas, sobre los cuales se tienen 2 líneas con densidades estimadas que se ajustan a los histogramas. Dicha gráfica se observa en la figura 1.4 en la cual podemos ver que las barras rojas representan el número de alumnos del turno vespertino y las barras azules representan el número de alumnos del turno matutino.

Notamos que en este caso las densidades son completamente diferentes. Algunos datos que se pueden obtener de dichas densidades 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.

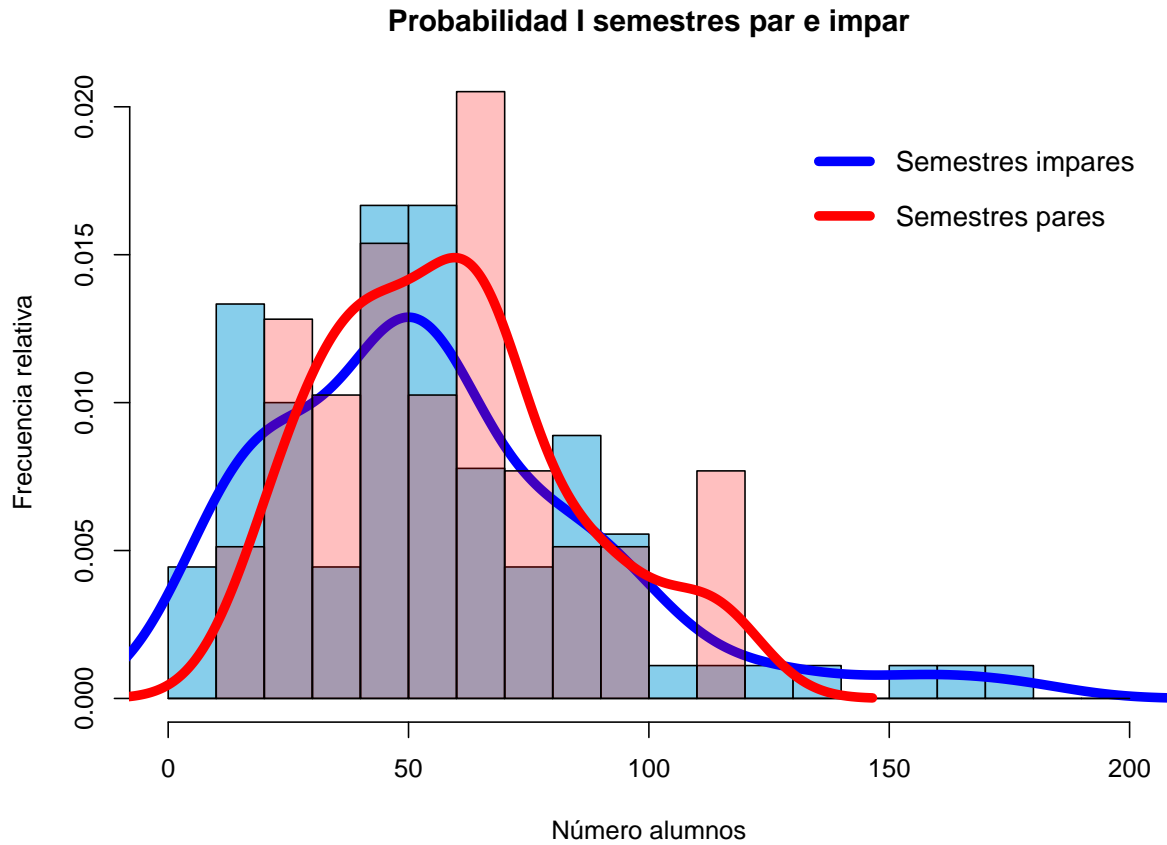


Figura 1.2: *Histograma del número de alumnos por semestre*

Con los resultados observados se obtuvieron los grupos de datos  $G_1, G_2, G_3, G_4$ , para hacer los análisis estadísticos, los cuales se definen en la tabla 1.2.

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

Tabla 1.2: *Grupos de datos*

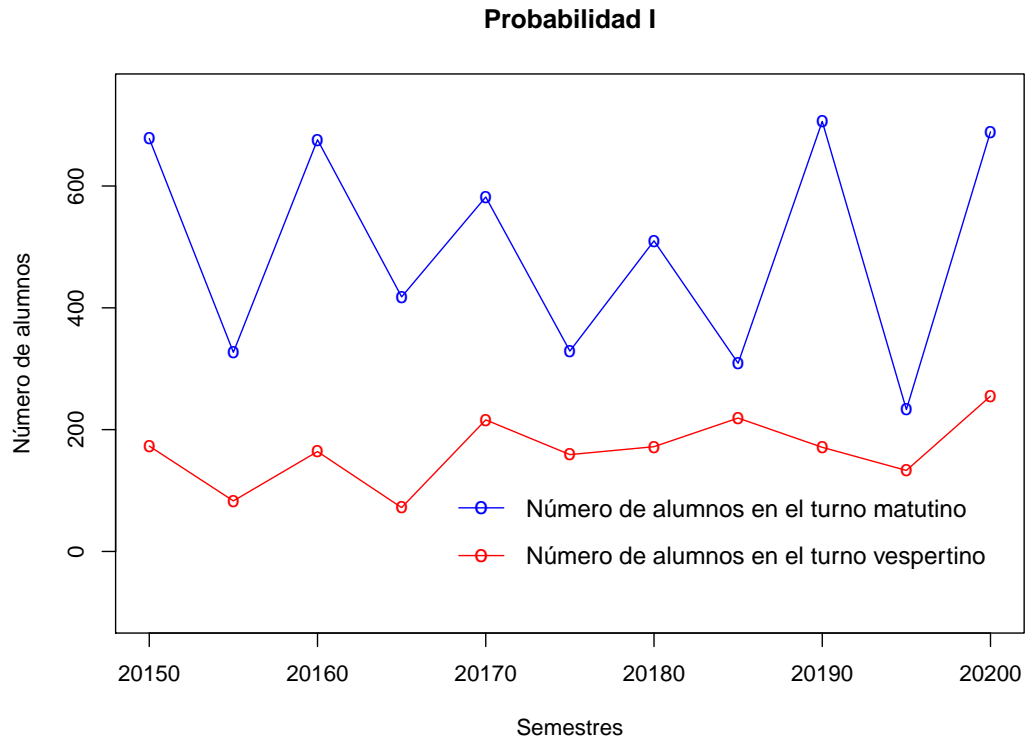


Figura 1.3: *Número de alumnos por turno*

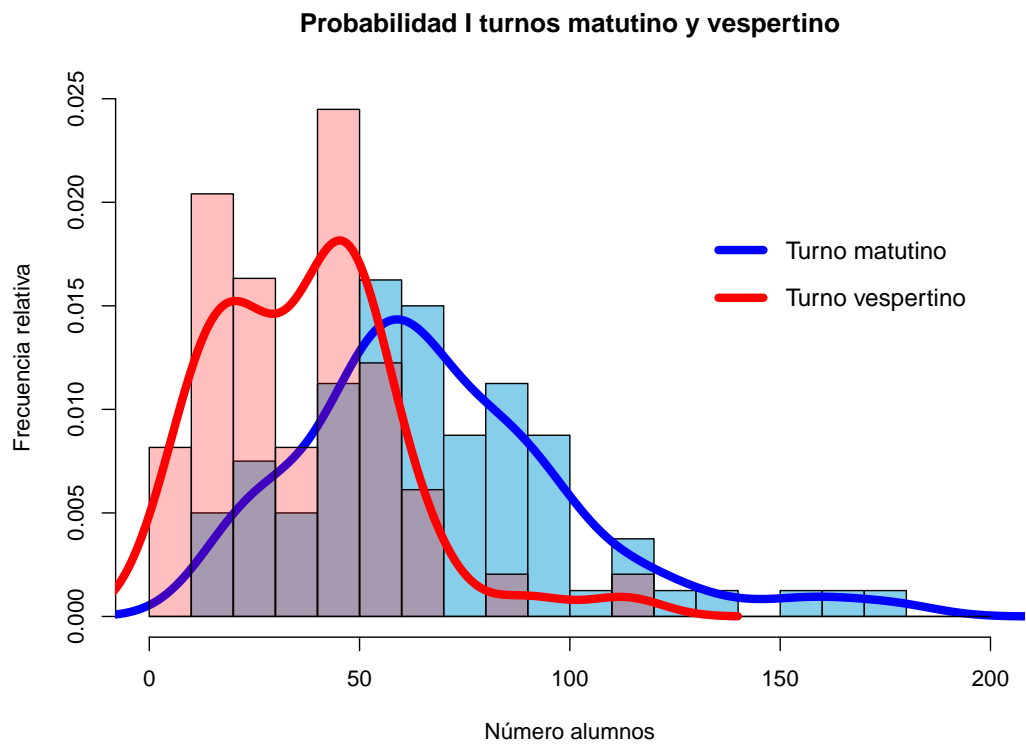


Figura 1.4: *Histograma del número de alumnos por turno*

# Capítulo 2

## Extracción de datos

La fuente de información de donde se obtuvieron los datos utilizados son las páginas de los horarios de la Facultad de Ciencias de la UNAM. Cada página contiene toda la posible información de los grupos de una materia, un semestre y una carrera. Sólo se toma en cuenta la información de las carreras del Departamento de Matemáticas, las cuales son: Actuaría, Ciencias de la Computación, Matemáticas y Matemáticas Aplicadas.

La información que se puede extraer de las páginas mencionadas es: nombre de profesores, nombre de ayudantes, salón, horario, plan, carrera, año y tipo de semestre (ej. *2018-2* implica que es el semestre par del 2018), materia, semestre de la materia. (ej. *Quinto Semestre*), tipo de materia (obligatoria u optativa), fecha, horario y salón de exámenes finales. En la figura 2.1 se muestra un ejemplo de dichas páginas.

The screenshot shows the UNAM website's 'Horarios 2020-1' section for the Faculty of Sciences. The navigation bar includes links like Inicio, Contacto, Mapa del sitio, Directorio, Correo, Tienda Virtual, and Ingresar. The main content area is titled 'Matemáticas (plan 1983)' and lists 'Lenguajes de Programación y sus Paradigmas, Optativas de los Niveles V y VI'. It features a search bar and a 'Buscar en grupos' button. Two groups are listed: Grupo 7068 (40 places, 37 students) and Grupo 7070 (40 places, 31 students). Each group has a section for 'Exámenes finales' with dates and a list of faculty members and their schedules.

Grupo	Lugares	Alumnos	Exámenes finales	Profesores	Horarios
Grupo 7068	40	37	viernes 29 de noviembre 2019 y viernes 6 de diciembre 2019 de 11 a 13 en el O125.	Profesor Favio Ezequiel Miranda Perea	lu mi vi 11 a 12 O125
				Ayudante Javier Enríquez Mendoza	ma ju 11 a 12 O125
				Ayud. Lab. Pablo Gerardo González López	mi 14 a 16 Laboratorio de Ciencias de la Computación 2
Grupo 7070	40	31	martes 26 de noviembre 2019 y martes 3 de diciembre 2019 de 11 a 13 en el P211.	Profesor Fernando Abigail Galicia Mendoza	lu mi vi 11 a 12 P211
				Ayudante Pedro Juan Salvador Sánchez Pérez	ma ju 11 a 12 P211
				Ayud. Lab. María Ximena Lezama Hernández	mi 14 a 16 Laboratorio de Ciencias de la Computación 1

Figura 2.1: Página de horarios de la FC

## 2.1. Extracción de información con la aplicación Selector-Gadget

Para extraer los datos de las páginas de la FC se utilizó una aplicación de Google Chrome llamada *SelectorGadget*, la cual permite seleccionar la información deseada y ésta arroja una sección del código CSS de la página web el cual se introduce en R para que se descargue la información deseada.

En la figura 2.2 se puede ver un ejemplo del funcionamiento de la aplicación. El ícono de la aplicación es una lupa, el cual se encuentra señalado por la flecha roja. Una vez presionado el ícono se seleccionó el nombre de algún profesor (en color verde), la aplicación automáticamente seleccionó todas las entradas que coincidían (en color amarillo), pero debido a que se seleccionó más información de la deseada entonces se dio click sobre un salón y el nombre de un ayudante (en color rojo). En el cuadro de texto, la aplicación arroja la sección del código CSS correspondiente, junto al código mencionado, se puede observar que se seleccionaron 15 entradas correspondientes a los nombres de los profesores de la materia de Probabilidad I.

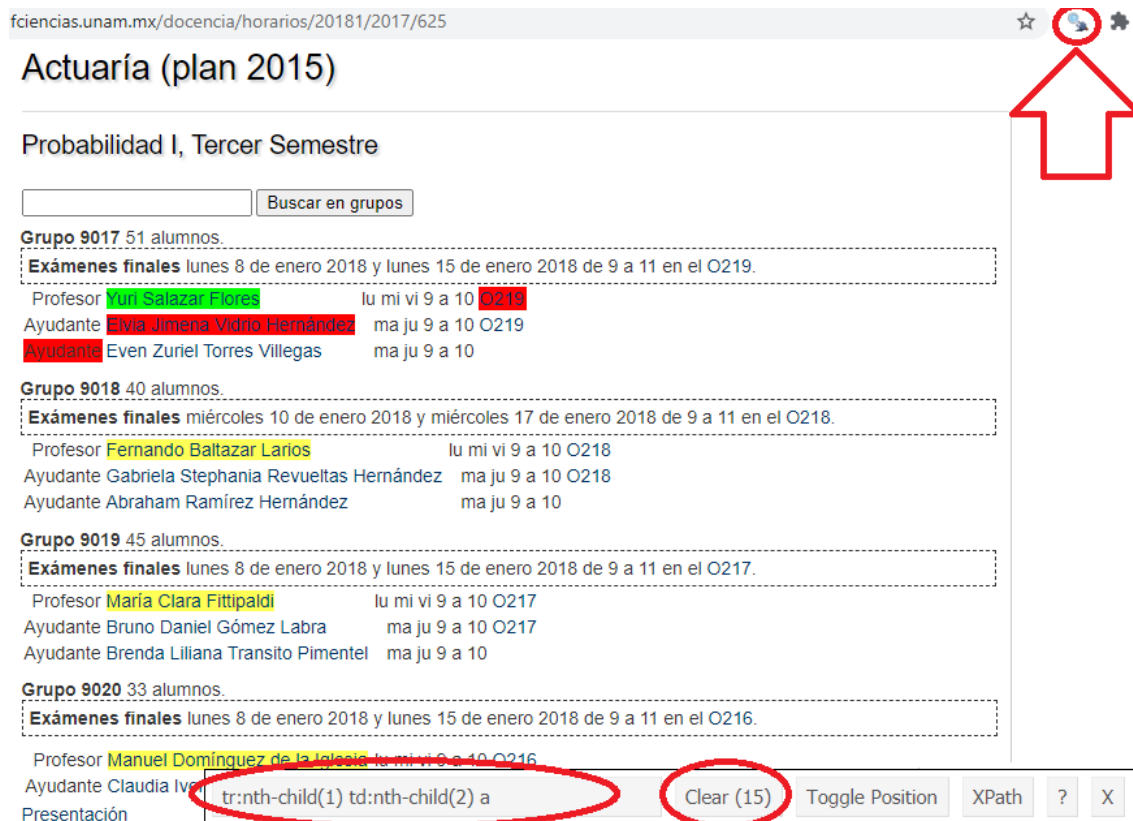


Figura 2.2: Aplicación SelectorGadget

## 2.2. Estructura de las páginas web

El estar haciendo la búsqueda de información, de manera manual, de los horarios para cada materia, cada carrera y cada uno de sus planes de estudio, requiere de mucho tiempo de

trabajo, por lo que decidimos buscar la existencia de alguna estructura en las páginas web de los horarios de la FC para poder realizar la búsqueda de la información de una manera automática y mucho más rápida. Se observó que la estructura que siguen todas las páginas de internet 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 tipo de semestre (par o impar),  $a \in \{20081, 20082, 20091, 20092, 20101, 20102, \dots, 20201\}$

b = clave de plan de estudio

c = número de materia,  $c \in \{1, 1000\}$ .

Para la realización de este trabajo se tomaron en cuenta sólo los planes de estudio vigentes al día de hoy, los cuales son todos los planes mostrados en la tabla 2.1, salvo el plan 1972 de Actuaría. 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: Planes de estudio por carrera con clave

Una vez identificada la estructura de las páginas web se pudo realizar la búsqueda de información de manera automatizada. Se decidió que  $c \in \{1, 2, 3, \dots, 10000\}$ .

Se hizo una función que genera una matriz llamada *mat\_posibles\_url*, en la cual se guardan, entre otros datos, las URL's de las páginas de los horarios de la FC. La función sólo guarda las URL's que si existen.

Al obtener la matriz, se observó que el valor máximo que toma  $c$  es 991, por lo que se redujo su conjunto de posibles valores y se definió  $c \in \{1, \dots, 1000\}$ .

Las categorías de las columnas de la matriz *mat\_posibles\_url* se definen en la tabla 2.2.



Col.	Nombre	Explicación	Posibles valores
1	Semestre	Semestre al que pertenece la materia (Año y semestre)	1 <sup>o</sup> , 2 <sup>o</sup> , ..., 8 <sup>o</sup>
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 FC	Páginas web de FC
5	Num. Grupos	Número de grupos que hay en cada página de internet	N

Tabla 2.2: Descripción de columnas de *mat\_posibles\_url*

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

Al inicio se encontraron tres tipos de grupos dentro de las páginas de horarios de la FC, cada uno con información similar, pero se hizo la separación de acuerdo a sus diferencias.

En este trabajo se considera como semestre actual al semestre 2020 – 1.

En todos los grupos se puede encontrar la información del nombre de profesor, nombre de ayudante, salón, horario y el número de alumnos inscritos en el grupo.

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

**Grupo 9301, 129 lugares. 84 alumnos.**

Profesor José Luis Navarro Urrutia	lu mi vi 13 a 14 Aula Magna I
Ayudante Luz Candy Becerril Palacios	ma ju 13 a 14 Aula Magna I
Ayudante Gabriela Yaneth Romo Cordoba	ma ju 13 a 14
Ayudante Adrián Gallardo Pacheco	ma ju 13 a 14

Figura 2.3: Tipo de grupo A

- b) En el grupo **B** se tienen las páginas correspondientes a semestres entre el 2018 – 2 y el semestre anterior al actual, con respecto al año en curso. En este tipo de grupos se tiene información del número de lugares disponibles por salón y la información de los exámenes finales.

**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: Tipo de grupo B

- c) En el grupo 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.

**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: Tipo de grupo C

## 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, ya que es en ese momento en donde se encuentran los diferentes problemas que se pueden presentar, como posibles errores en los datos, información incompleta, o valores poco comunes de acuerdo al comportamiento observado.

En nuestro caso al limpiar los datos nos encontramos con problemas de falta de información, problemas de información repetida, problemas al extraer información, entre otros, cada uno de éstos se desglosan en las siguientes secciones.

### 2.4.1. Problemas de falta de información

Se encontraron diferentes tipos de páginas que tenían grupos sin información e incluso páginas sin información alguna. A continuación se muestran varios ejemplos con los diferentes casos encontrados.

- 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>

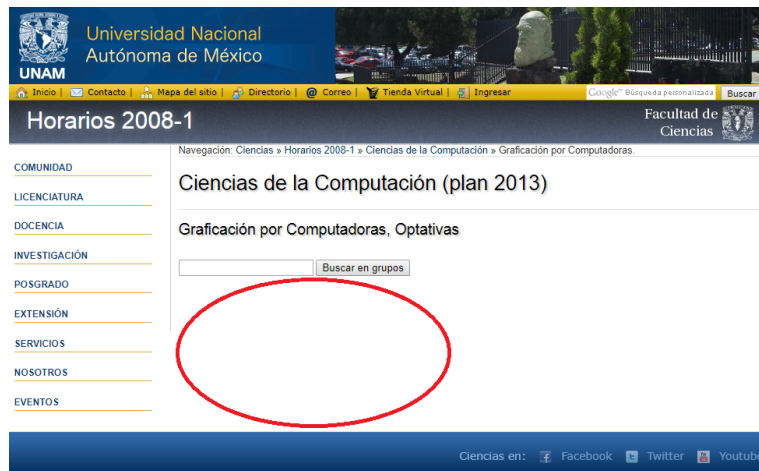


Figura 2.6: Ejemplo de página web en blanco

- 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: Ejemplo de grupo sin información de salón

- Páginas que tienen grupos sin información del número de alumnos inscritos en el grupo: <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: Ejemplo de grupo sin información de alumnos

- Páginas que tienen grupos sólo con el horario, sin nombre del profesor, salón, ayudante, número de alumnos, lugares disponibles: <http://www.fciencias.unam.mx/docencia/horarios/20091/119/841> <http://www.fciencias.unam.mx/docencia/horarios/20091/119/244>

## 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: Ejemplo de grupo sólo con horario

### 2.4.2. Problemas de información repetida

Dentro de los problemas de información repetida, se encontraron los siguientes casos:

- Tener información de una materia correspondiente a un plan de estudios posterior al semestre en el que se está buscando la información: <http://www.fciencias.unam.mx/>

docencia/horarios/20082/1556/803 y tener la misma información con el plan de estudios correspondiente: <http://www.fciencias.unam.mx/docencia/horarios/20082/218/803>

El número del plan de estudios corresponde al año en que entró en vigencia el plan. En la siguiente figura se puede ver una materia de la carrera de Ciencias de la Computación del semestre 2008-2, con planes distintos.

**Horarios 2008-2**

**Ciencias de la Computación (plan 2013)**

Graficación por Computadoras, Optativas

**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

**Horarios 2008-2**

**Ciencias de la Computación (plan 1994)**

Graficación por Computadoras, Optativas

**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
(b) Plan de estudios correspondiente

Figura 2.10: Ejemplo de información repetida: Planes de estudio

- Tener una misma materia con nombres distintos para las diferentes carreras: <http://www.fciencias.unam.mx/docencia/horarios/20201/217/1712> para Matemáticas, plan 1983 y <http://www.fciencias.unam.mx/docencia/horarios/20201/2017/1739> para Actuaría, plan 2015. Notamos que la información en ambas páginas es la misma, sólo se cambian las claves de los grupos.

**Matemáticas (plan 1983)**

**Estadística III, Optativas de los Niveles VII y VIII**

**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

**Actuaría (plan 2015)**

**Modelos de Supervivencia y de Series de Tiempo, Séptimo Semestre**

**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

(a) Matemáticas: 1983
(b) Actuaría: 2015

Figura 2.11: Ejemplo de información repetida: Materia con nombres distintos

- Profesores que imparten dos o más clases distintas en el mismo horario y diferente salón: <http://www.fciencias.unam.mx/docencia/horarios/20111/2017/162> para Ecuaciones Diferenciales I y <http://www.fciencias.unam.mx/docencia/horarios/20111/2017/91> para Cálculo Diferencial e Integral I.

Las materias mencionadas son diferentes, pero las clases comienzan a la misma hora, Ecuaciones de 18-19hrs y Cálculo 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 4172** 15 alumnos.

Profesor Edgar René Hernández Martínez lu mi vi 18 a 19 O123

Ayudante Norma Angélica Cruz Cervantes ma ju 18 a 19 O123

(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: *Ejemplo de información repetida: Mismo profesor, materias distintas*

### 2.4.3. Otros problemas al extraer información

Al extraer la información surgieron otros problemas, en algunos casos se tuvieron que analizar las materias a mano. 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.

**Grupo 6125** *Un* alumno.

Profesor Reyna Pineda González lu mi vi 21 a 22 102

Ayudante Elmo Jesús Vilorio López ma ju 21 a 22 102

Figura 2.13: *Ejemplo de grupo con un alumno*

Para resolver este problema se identificó la variable tipo *string* igual a *Un* para convertir la información y que los datos obtenidos pudieran ser utilizados.

- El algoritmo supone que todas las clases duran un hora y no se consideran las medias horas: <http://www.fciencias.unam.mx/docencia/horarios/20172/1556/820>  
Se considera que esta 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: Ejemplo de grupo con medias horas

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

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

**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: Ejemplo de grupo con horarios múltiples

- Las materias de inglés no se imparten todos los días de la semana, en algunos casos se imparten clases en línea: <http://www.fciencias.unam.mx/docencia/horarios/20202/2017/1135>. Se registran únicamente los horarios de los días en que se imparten las clases presenciales.

**Grupo 9296**, 45 lugares. 20 alumnos.  
 Profesor Lillian Moreno Roldán sá 7 a 9 Sesión virtual  
 ma 14 a 16 P207

Figura 2.16: Ejemplo de grupo de inglés

- Se tienen grupos que no tienen la misma estructura que los tipos de grupos **A**, **B** y **C** definidos en la sección 2.3: <http://www.fciencias.unam.mx/docencia/horarios/>



20201/2017/872, debido a ello el código CSS utilizado no sirve para obtener toda la información que se puede obtener del grupo.

En este caso no se lee adecuadamente el número de alumnos inscritos en el grupo.

## 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: Ejemplo de grupo con estructura diferente

## 2.5. Matrices de datos

Una vez que se realizó el proceso de la 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:

Col.	Nombre	Explicación	Posibles valores
1	Materia	Nombre del curso impartido	“Probabilidad I”
2	Profesor	Nombre de la persona que va a impartir alguna materia	“Arrigo Coen Coria”
3	Horario	Hora en la que se imparte alguna materia	“7 a 8”, ... , “21 a 22”
4	horario_num	Valores de la columna Horario en variables tipo <i>numeric</i>	7, 8, 9, ..., 20, 21
5	Lugares	Espacios disponibles por salón	$\mathbb{N}$
6	Alumnos	Número de estudiantes inscritos por grupo	$\mathbb{N}$
7	Salón	Espacio físico en el que se imparte alguna materia	“O123”, ..., “P105”
8	Grupo	Clave con la que se identifica una asignación	4489, 6114, ...
9	Carrera	Nombre de alguna carrera de FC	“Actuaría”, “Mate-máticas”,...
10	Plan	Año en el que se implemento un nuevo plan de estudios	1983, ..., 2017
11	Semestre	Semestre al que pertenece la materia (Año y semestre par o impar)	20081, ... , 20192, 20201
12	Cambios	Clave que indica los cambios que se le han hecho al grupo	$\mathbb{N}$

La tabla continúa en la siguiente página



Col.	Nombre	Explicación	Posibles valores
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”, ... , “Optativas”
15	url	Nombre de la página de los horarios de FC correspondiente al grupo	url's de FC
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
23	NomMat_Act2000	Indica el nombre de las materia correspondiente a la carrera de Actuaría plan 2000	Nombres de materias de FC
24	NomMat_Act2006	Indica el nombre de las materia correspondiente a la carrera de Actuaría plan 2006	Nombres de materias de FC
25	NomMat_Act2015	Indica el nombre de las materia correspondiente a la carrera de Actuaría plan 2015	Nombres de materias de FC
26	NomMat_CdC1994	Indica el nombre de las materia correspondiente a la carrera de CdC plan 1994	Nombres de materias de FC
27	NomMat_CdC2013	Indica el nombre de las materia correspondiente a la carrera de CdC plan 2013	Posibles valores
28	NomMat_Mat1983	Indica el nombre de las materia correspondiente a la carrera de Matemáticas plan 1983	Nombres de materias de FC
29	NomMat_MAp2017	Indica el nombre de las materia correspondiente a la carrera de MatAp plan 2017	Nombres de materias de FC
30	URL_Act2000	Indica la URL correspondiente a la carrera de Actuaría plan 2000	url de FC
31	URL_Act2006	Indica la URL correspondiente a la carrera de Actuaría plan 2006	url de FC
32	URL_Act2015	Indica la URL correspondiente a la carrera de Actuaría plan 2015	url de FC
33	URL_CdC1994	Indica la URL correspondiente a la carrera de CdC plan 1994	url de FC
34	URL_CdC2013	Indica la URL correspondiente a la carrera de CdC plan 2013	url de FC

La tabla continúa en la siguiente página

Col.	Nombre	Explicación	Posibles valores
35	URL_Mat1983	Indica la URL correspondiente a la carrera de Matemáticas plan 1983	url de FC
36	URL_MAp2017	Indica la URL correspondiente a la carrera de MatAp plan 2017	url de FC
37	Num_materia	Número del nombre de materia de acuerdo al vector <i>vec_nom_materias</i>	$\mathbb{N}$

Tabla 2.3: Descripción de columnas de *m\_grande*

La columna *Cambios*, va a guardar todos los cambios que han “sobrevivido” esos grupos. El significado de los números que pueden aparecer en esa columna se explican a continuación:

- (1) Grupos revisados a mano.
- (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.
- (4) Páginas que no tienen información del salón.



# Capítulo 3

## Análisis estadístico

Debido a que los datos obtenidos no son independientes entre sí, 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, en donde los datos al tiempo presente dependen de las observaciones anteriores, es decir existe una dependencia de  $X_t$  con  $\{X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_1\}$ .

Denotamos a una serie de tiempo como:

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

Cada elemento de la serie de tiempo es llamado componente. A continuación se describe cada una de ellas.

- 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 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 nos indican, en su libro *The Analysis of Time Series An Introduction with R* (2), 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: Se tiene variación estacional aditiva y se utiliza cuando la varianza o la desviación estándar de la serie de tiempo se mantienen constantes a lo largo del tiempo.

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

2. Multiplicativo: 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.

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

3. Mixto: Se utiliza cuando se tiene variación estacional multiplicativa pero la variabilidad de la componente irregular se mantiene constante a lo largo del tiempo.

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

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 periodos de información para que la predicción sea lo más acertada posible.

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

### 3.1. Análisis estadístico básico

En esta sección se hará un análisis de los datos obtenidos. 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 se va a utilizar para mostrar dichos patrones y para realizar predicciones de los datos es el método Holt-Winters aditivo.

El método se utiliza para describir y predecir valores con series de tiempo que tienen componentes de tendencia y de estacionalidad. Existen pruebas, que hemos hecho a los datos, para comprobar que los datos que obtuvimos cumplen con los supuestos del método. Éstas pruebas se muestran a lo largo de esta sección.

Primero se van a graficar los datos como serie de tiempo y después se van a mostrar algunas gráficas de barras. Vamos a observar el comportamiento de los datos, ver si tiene alguna tendencia y variación estacional.

En la figura 3.1 se muestra la gráfica de barras con el número total de alumnos que toman clases por semestre. A simple vista notamos que tiene una tendencia creciente y una estacionalidad semestral. Podemos ver también que el número de alumnos de los semestres impares es siempre mayor al de los semestres pares. Éste fenómeno los vimos en la figura 1.1 al hacer el análisis correspondiente a los datos de la materia de Probabilidad I.

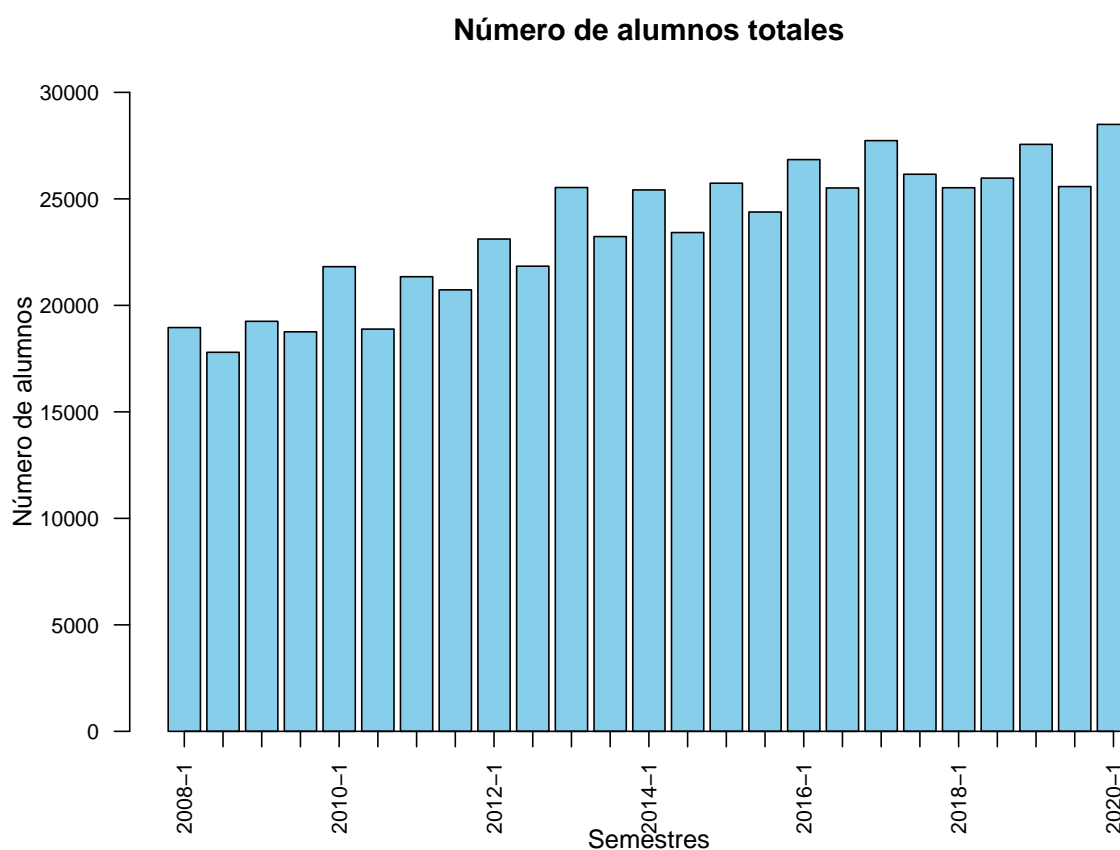


Figura 3.1: *Número total de alumnos por semestre*

En la figura 3.2 se muestra la gráfica de la media del número de alumnos que toman clases por semestre de todas las materias. Observamos que los valores tienen una tendencia creciente, esto nos indica que cada semestre, en promedio hay más alumnos tomando clases en la FC.



Figura 3.2: *Media de alumnos por semestre: Se observa una tendencia creciente*

En la figura 3.3 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. Su rango se encuentra entre 24 y 29.



Figura 3.3: *Desviación estándar del número de alumnos por semestre*

En la figura 3.4 se observan 4 diferentes gráficas, en la primera se observan los datos reales. En la segunda la tendencia de los datos, la cual notamos que es creciente. En la tercera la componente estacional que nos indica que los datos tienen una estacionalidad semestral. En la cuarta se ve la componente aleatoria de los datos la cual ya no tiene estacionalidad ni tendencia.



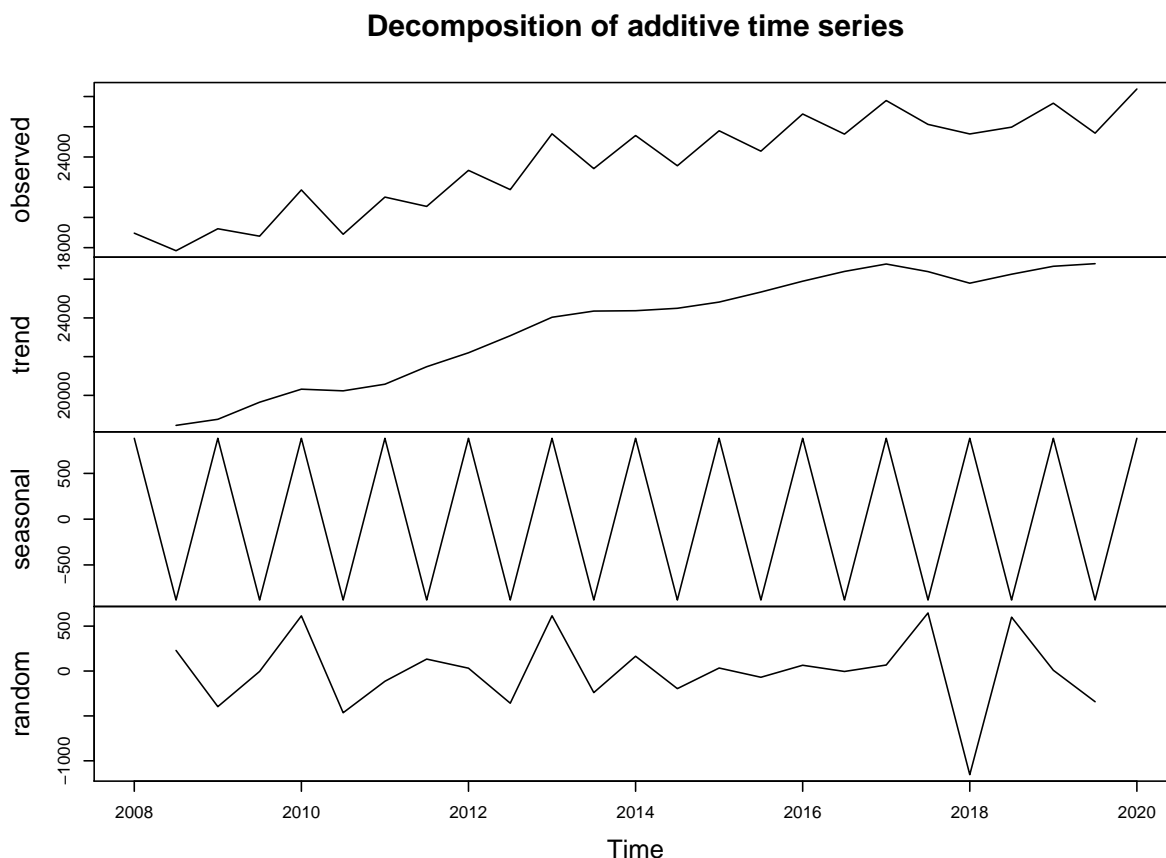


Figura 3.4: *Descomposición por el método aditivo de Holt-Winters: Total de alumnos por semestre*

Observando la figura 3.4 notamos que los datos tienen estacionalidad semestral y una tendencia creciente, por lo tanto confirmamos que se puede utilizar el método Holt-Winters ya que se cumplen los supuestos que se requieren. Para verificar que el modelo de estacionalidad adecuado es el aditivo, vemos la figura 3.3 y notamos que la desviación estándar permanece constante a lo largo del tiempo.

## 3.2. Análisis estadístico por grupo de datos

En la figura 3.5 se muestra la gráfica del número de alumnos separado por semestres pares e impares. Se observa un comportamiento similar al de la figura 1.1. Vemos con mayor claridad lo que ocurre en la figura 3.1, los datos efectivamente tienen una tendencia creciente. Sin embargo notamos que el número de alumnos de los semestres impares es mayor al número total de alumnos de los semestres pares, salvo en el semestre 2018-1 en donde el número de alumnos es menor a los de los semestres adyacentes, los cuales son el 2017-2 y el 2018-2.

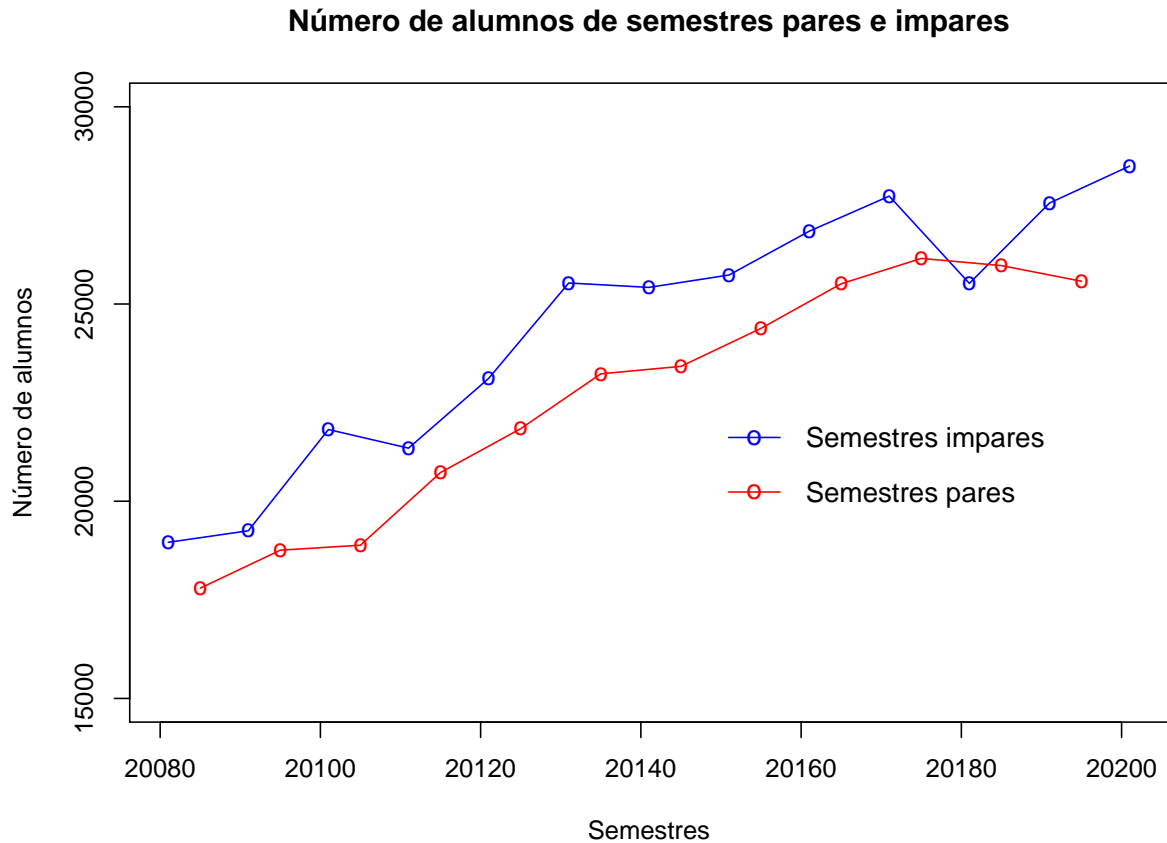


Figura 3.5: *Número de alumnos de semestres pares e impares*

En la figura 3.6 se muestra el histograma doble del 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. 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 la distribución que mejor se le ajusta a los datos es la Poisson, así como lo vimos en la figura 3.9.

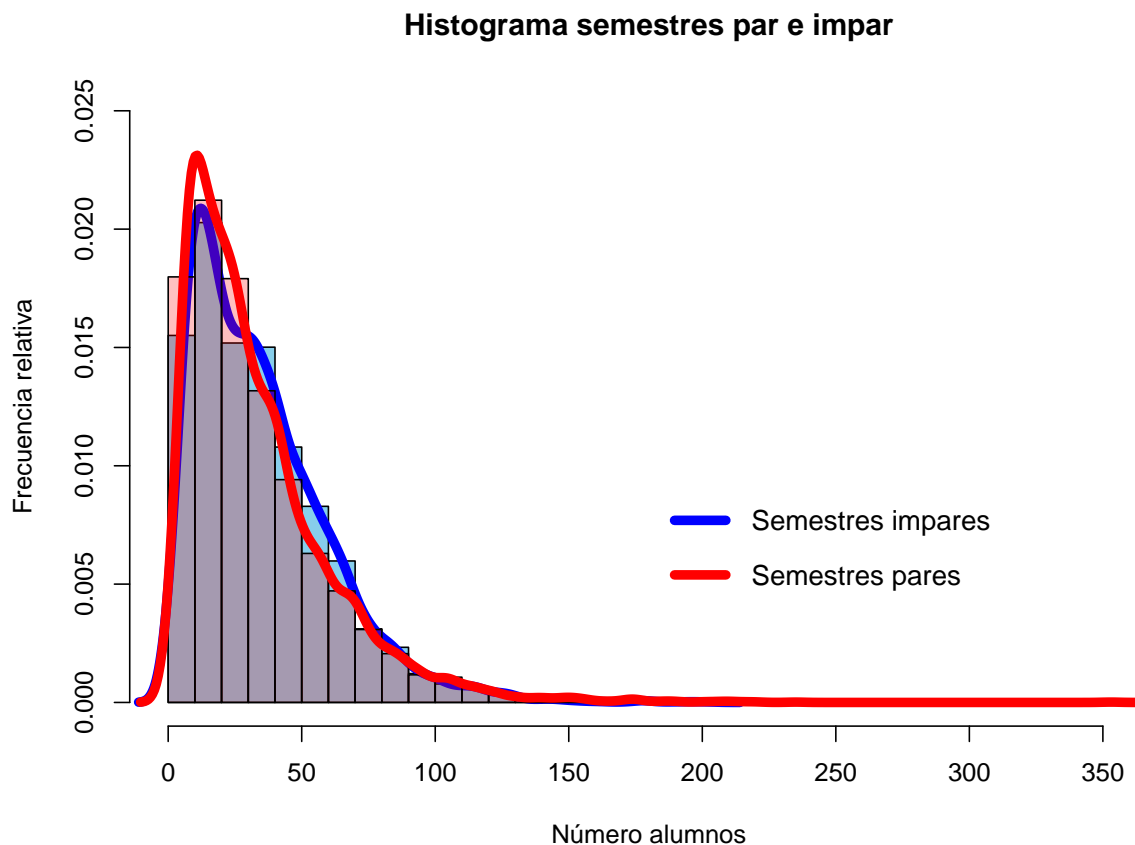


Figura 3.6: *Histograma del número de alumnos de semestres pares e impares*

En la figura 3.7 se muestra 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. Este comportamiento lo observamos en la figura 1.3 al hacer el análisis para la materia de *Probabilidad I*.

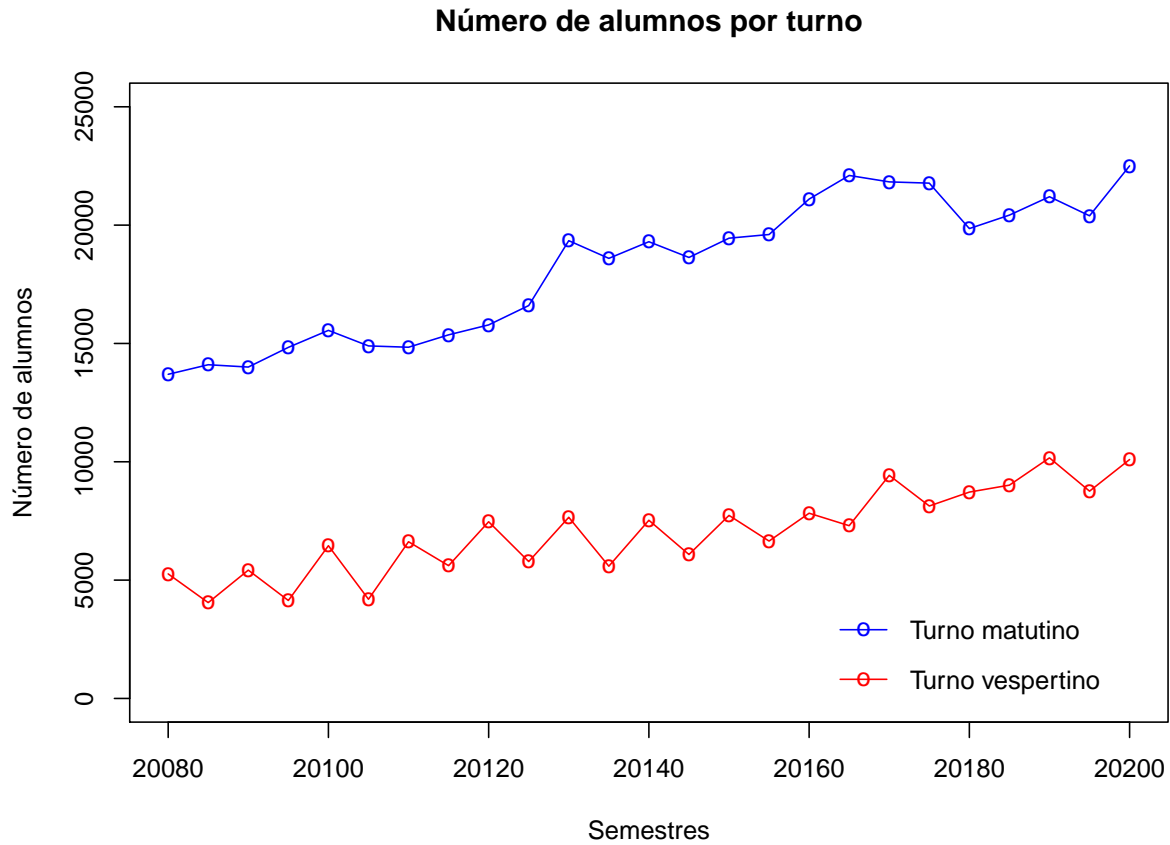


Figura 3.7: *Número de alumnos por turno de todos los semestres*

Los datos que se graficaron en el histograma de la figura 3.8 son los alumnos totales por hora de cada semestre. En dicha figura se muestra el histograma doble de los datos divididos en los turnos matutino y vespertino. Notamos que la diferencia entre cada turno es evidente. Al ver la gráfica podemos concluir que hay más alumnos en el turno matutino que en el vespertino.

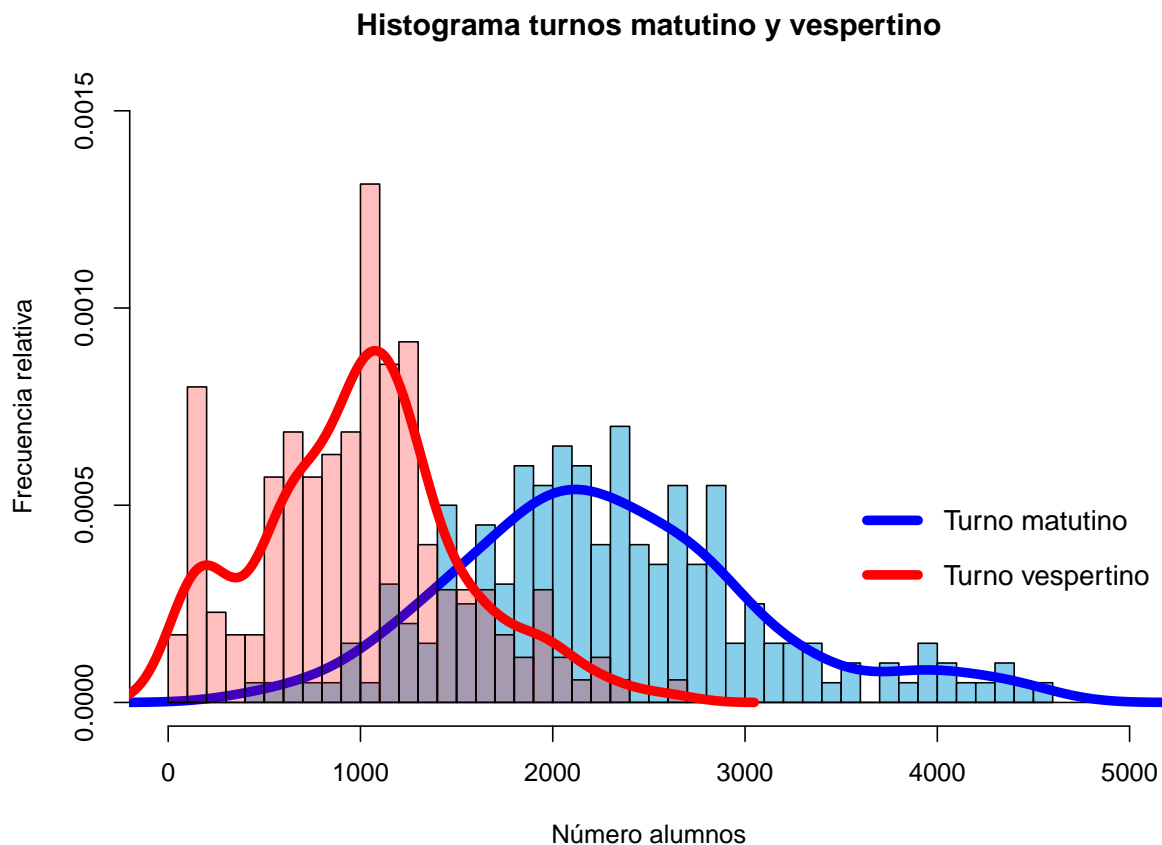


Figura 3.8: *Histograma del número de alumnos de los turnos matutino y vespertino*

### 3.3. Distribución del tamaño de los grupos

En la figura 3.9 se muestra el histograma del número de alumnos por grupo de todos los semestres.

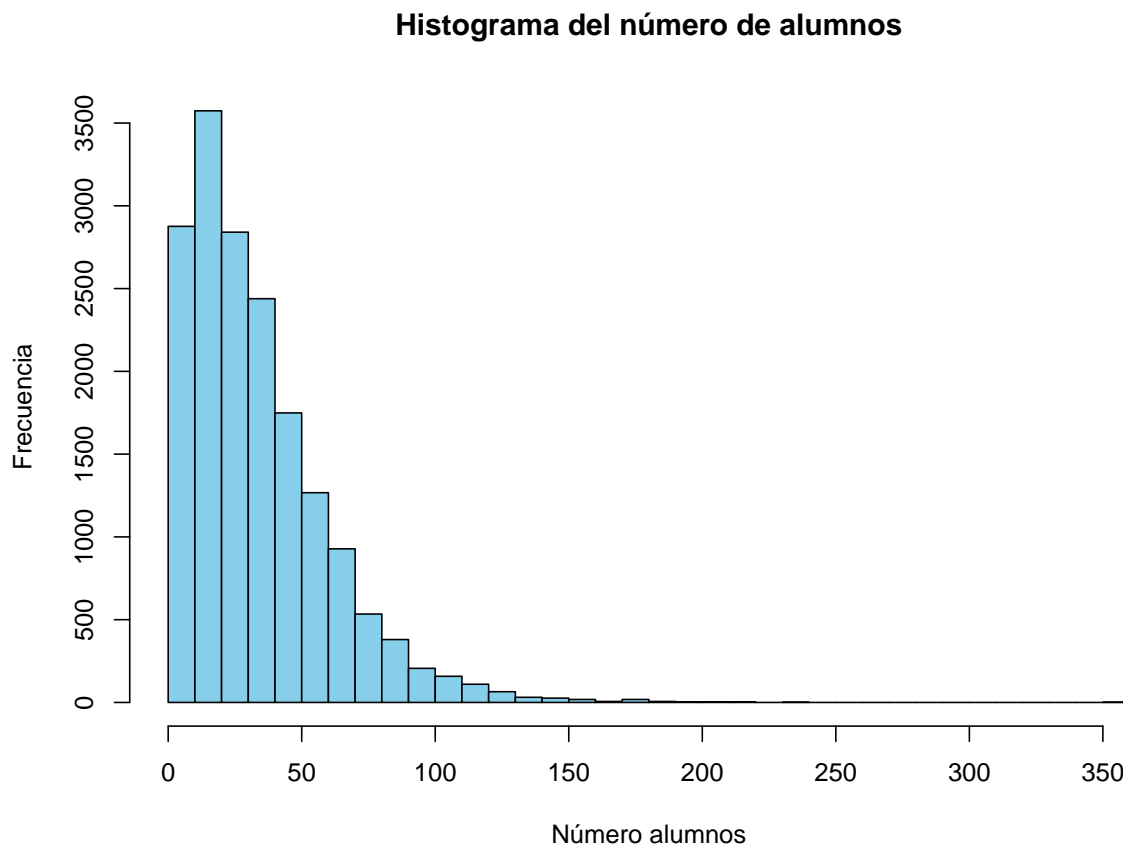


Figura 3.9: *Histograma del número de alumnos por grupo de todos los semestres*

En la figura 3.10 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. Notamos que el comportamiento es prácticamente el mismo en todos los semestres, salvo en dos cuyo pico máximo es mucho menor que los demás.

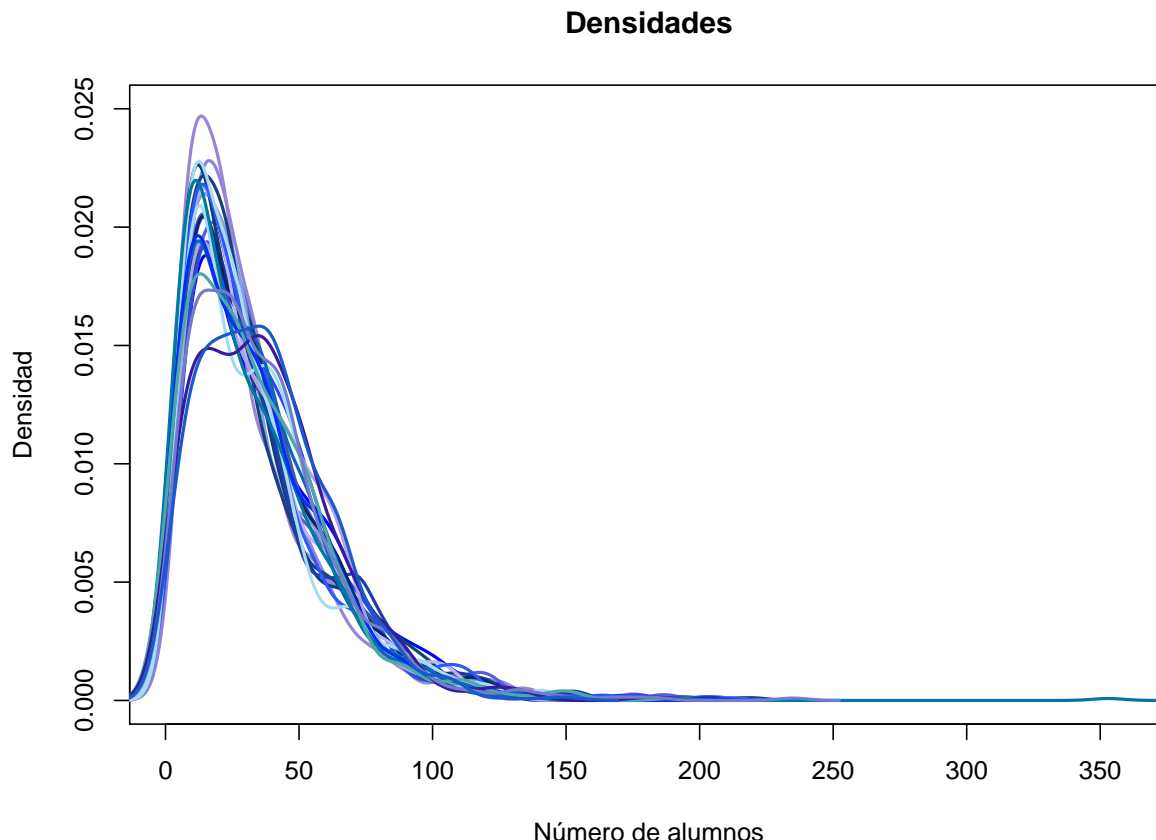


Figura 3.10: *Densidades del número de alumnos por grupo de cada semestre*

Viendo las gráficas de las figuras 3.9 y 3.10, 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. La prueba de esta hipótesis la veremos en la sección 4.4.

### 3.4. 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 3.11 se muestra la gráfica de barras con el número promedio de grupos por hora. Se tomó la información de 25 semestres. Se observa una disminución considerable de los 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 FC.

Hay un descenso leve a las 9hrs donde se pudiera suponer que la gente sale a desayunar. En mi caso particular, al estudiar la carrera de Actuaría, tenía clases desde las 7a.m. por lo que las 9a.m. era una buena hora para hacer una pequeña pausa en el horario y tener un descanso

para desayunar. Esta hora coincide con el cambio en el que se dejan de impartir las materias exclusivas para los actuarios como Teoría del Seguro, MASP o MASD. Es decir, 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 menor número de grupos.

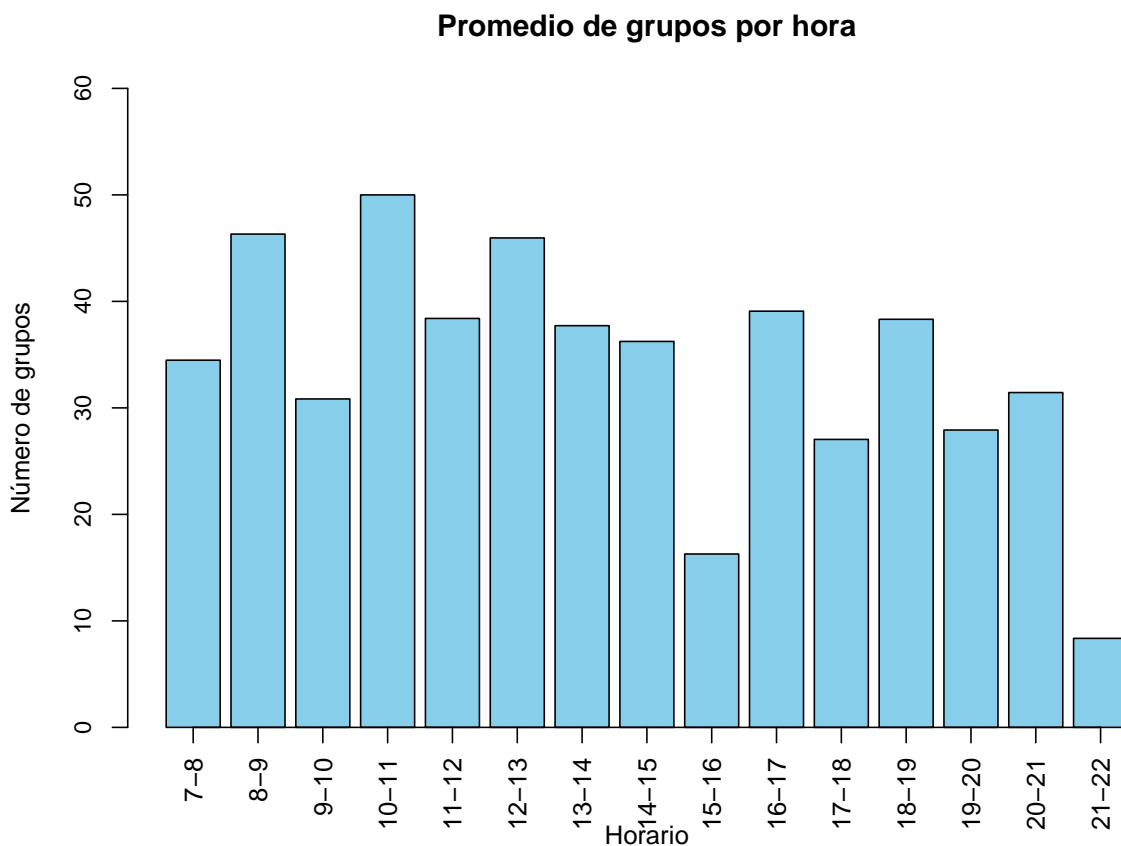


Figura 3.11: *Número promedio de grupos por hora*

En la figura 3.12 se muestra la gráfica de barras con el promedio número de alumnos por hora. Notamos que el comportamiento de ésta gráfica es muy similar al de la gráfica mostrada en la figura 3.11. 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. Esta correlación que existe entre ambos tipos de datos ...



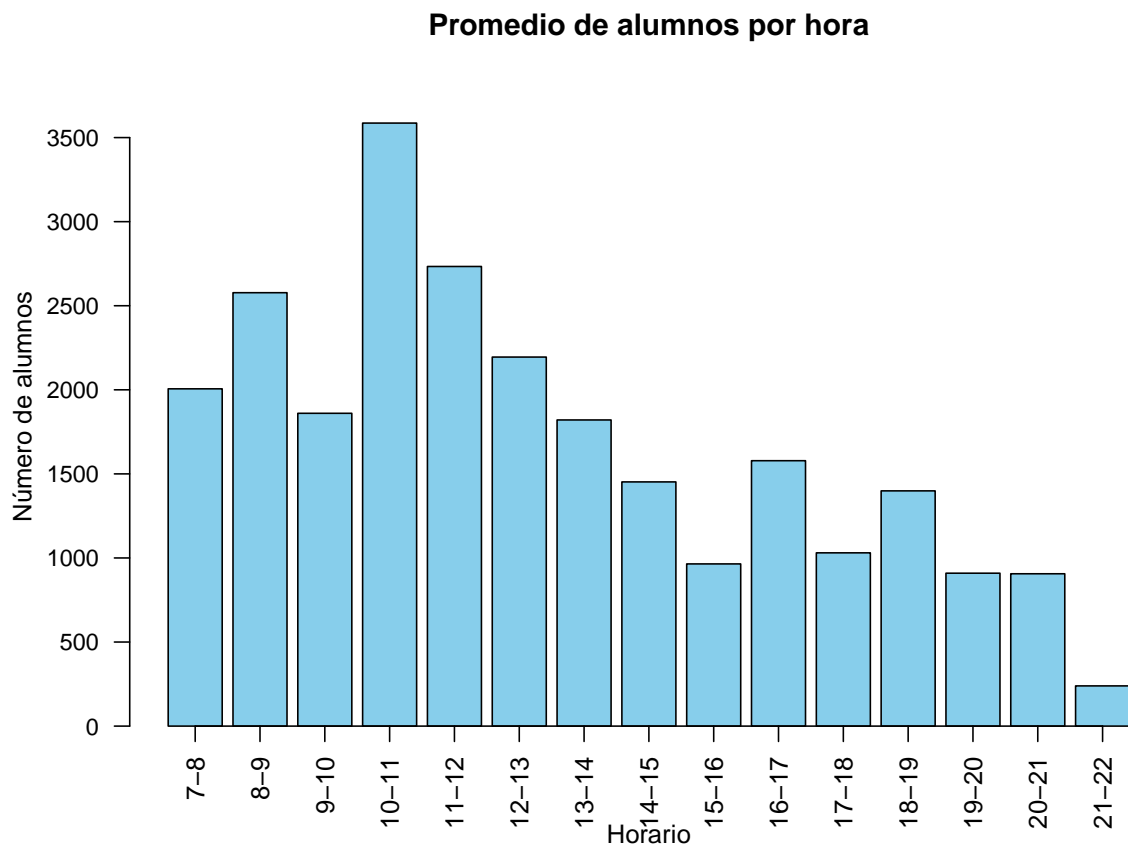


Figura 3.12: *Número promedio de alumnos por hora*

# Capítulo 4

## Simulación

La simulación es un proceso que nos permite estudiar el comportamiento de un sistema complejo que es muy difícil de examinar de manera analítica. Nos ayuda a determinar de manera empírica las probabilidades de ciertos eventos. La simulación nos permite experimentar con diversos supuestos que podrían ser muy costosos de realizar. Las áreas en las que se utiliza la simulación como herramienta diaria son por ejemplo en biología, estadística, medicina, química, matemáticas, investigación de operaciones, física, ingeniería o en las ciencias sociales.

Algunos 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. Se utiliza para todo tipo de propósitos, por ejemplo para poder realizar predicciones en base a datos históricos o enseñar a los pilotos a volar un avión sin poner en riesgo a la población al volar el avión real.

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

A lo largo de este capítulo explicaremos el proceso que seguimos para obtener la asignación de los horarios para cada materia con su respectivo profesor.

### 4.1. Funciones hechas en R

La función que genera esqueletos se llama *gen\_esqueleto* la cual carga la función *simula\_grupos* y ésta a su vez carga la función *estima\_grupos*.

El diagrama de flujo de la función se puede ver en la figura 4.1.

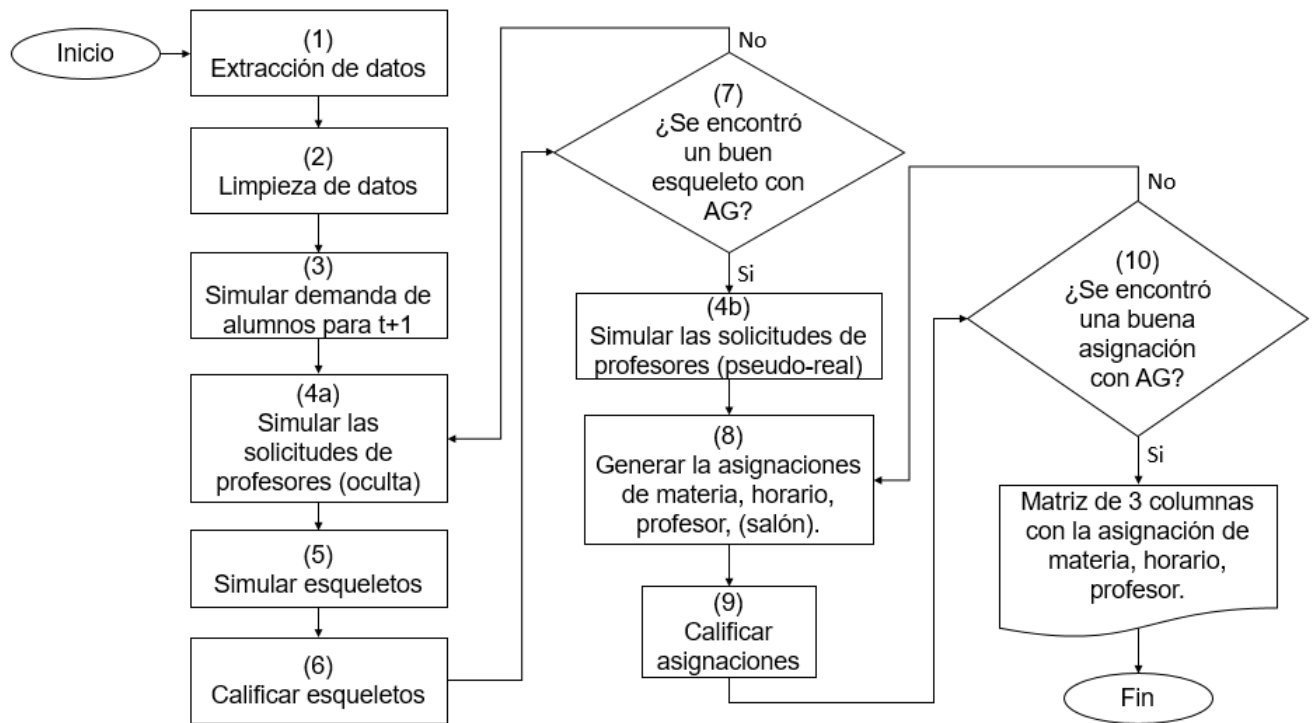


Figura 4.1: Diagrama de flujo de la función *gen\_asignacion*

- *posibles\_url*
- *gen\_m\_grande*
- *gen\_m\_grande\_total*
- *gen\_esqueleto*
- *gen\_solicitudes*
- *gen\_asignacion*
- *gen\_simula\_alumnos*
- *gen\_simula\_tamano\_grupo*

## 4.2. Obtención de los parámetros $q_1$ y $q_2$

En esta sección vamos a explicar cómo obtuvimos los valores de  $q_1$  y  $q_2$ . Son parámetros que se introducen en la función *hw()* de *R*. Representan los cuantiles utilizados al calcular los intervalos de confianza. Por ejemplo si  $q_1 = 80$  entonces se calcula el intervalo al 80% de confianza. Si se introducen a la función los dos parámetros entonces se calculan dos intervalos, uno al  $q_1$  % de confianza y el otro al  $q_2$  % de confianza.

Primero seleccionamos los parámetros generales necesarios para las simulaciones:

1. Fijamos la semilla con `set.seed(8654)`.

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 I*, *Modelos no Paramétricos y de Regresión*, *Administración de Riesgos Financieros* y *Seminario 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 de  $q_1$  y la letra  $U$  indica que se tomó la cota superior de  $q_2$ . Con estas cotas formamos intervalos de los cuales obtuvimos las simulaciones para los 3 diferentes semestres previamente elegidos.

$q_1 \backslash q_2$	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: Posibles valores para  $q_1$  y  $q_2$

Una vez hecha la simulación obtuvimos una tabla con 7 columnas: materia, intervalo, mín, media, máx, sd y seg. Donde en el renglón  $i$  se tienen los datos de la matriz de diferencias relativas de la  $i$ -ésima materia para cada intervalo de  $q_1$  y  $q_2$ . En la figura 4.2 vemos los primeros 10 renglones de la tabla obtenida.

Materia	Intervalo	Min	Media	Max	sd	Seg
Cálculo Diferencial e Integral I	L80,U85	-2.622222	-0.21911543	0.8627586	0.7287619	7.50
Demografía I	L80,U85	-1.985714	-0.09395821	0.8378378	0.4695739	7.53
Modelos no Paramétricos y de Regresión	L80,U85	-6.922222	-0.45848861	1.0000000	1.5742750	6.27
Administración de Riesgos Financieros	L80,U85	-1.816667	-0.03119518	0.6312500	0.3145900	3.08
Seminario de Investigación de Operaciones	L80,U85	-2.300000	-0.05121275	0.9384615	0.4202550	5.50
Cálculo Diferencial e Integral I	L80,U90	-2.588889	-0.25311699	0.7220690	0.7108189	8.34
Demografía I	L80,U90	-3.228571	-0.20226571	0.7270270	0.6654177	7.16
Modelos no Paramétricos y de Regresión	L80,U90	-6.744444	-0.48396359	1.0000000	1.6007042	5.94
Administración de Riesgos Financieros	L80,U90	-2.316667	-0.04418860	0.6375000	0.3924680	3.28
Seminario de Investigación de Operaciones	L80,U90	-2.233333	-0.05595981	0.9461538	0.4210018	5.49

Figura 4.2: Matriz con medidas de dispersión

Decidimos elegir  $q_1$  y  $q_2$  en base a la desviación estándar. Con la tabla 4.2 obtuvimos una

matriz de dos columnas que contine en su primer columna el intervalo y en la segunda el promedio de la desviación estándar para cada intervalo de las 5 materias. Los datos de dicha tabla los podemos ver en la figura 4.3.

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: Promedio de la desviación estándar: 5 materias, 12 pruebas

Los datos en la tabla 4.3 están ordenados de menor a mayor con respecto al promedio de la desviación estándar. Para la segunda prueba se eligieron los primeros 6 intervalos de dicha tabla. Se eligieron 10 materias: *Álgebra Lineal I*, *Álgebra Superior II*, *Algoritmos Genéticos*, *Análisis Matemático IV*, *Análisis Numérico*, *Teoría de la Medida I*, *Cálculo Diferencial e Integral IV*, *Graficas y Juegos*, *Inglés I* y *Matemáticas Actuariales para Seguro de Daños*. La tabla con el promedio de la desviación estandar 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: Promedio de la desviación estándar: 10 materias, 6 pruebas

Para la tercera prueba elegimos, de la tabla 4.4 los intervalos que tuvieran un promedio en la desviación estándar menor a 0.5. Se eligieron otras 10 materias: *Estadística III*, *Teoría del*

*Seguro, Programación Entera, Investigación de Operaciones, Geometría Moderna I, Geometría Analítica II, Lógica Matemática I, Cálculo Diferencial e Integral III, Estadística I y Bases 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: Promedio de la desviación estándar: 10 materias, 4 pruebas

Podemos ver que los valores de la tabla 4.5 son muy parecidos. Se hizo una prueba con los mismos intervalos pero con 5 materias que se dan en todos los semestres y además tienen muchos alumnos. La prueba se hizo para ver si había alguna diferencia en los datos y se pudiera elegir un sólo intervalo. Las materias que se eligieron para esta prueba fueron: *Geometría Analítica I, Cálculo Diferencial e Integral II, Finanzas I, Probabilidad II y Procesos Estocásticos I*. 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: Promedio de la desviación estándar: 5 materias, 4 pruebas

Analizando la información de las matrices de las figuras 4.5 y 4.6, decidimos elegir los valores de  $q_1 = 85$  y  $q_2 = 80$ . Por lo que el intervalo que buscamos estará formado por la cota inferior del intervalo de confianza al 85% y por la cota superior del intervalo de confianza al 80%. Para visualizar de una mejor manera cómo se encuentra el intervalo formado, podemos ver la figura 4.7. De dicho intervalo vamos a obtener los valores para simular la demanda de alumnos.

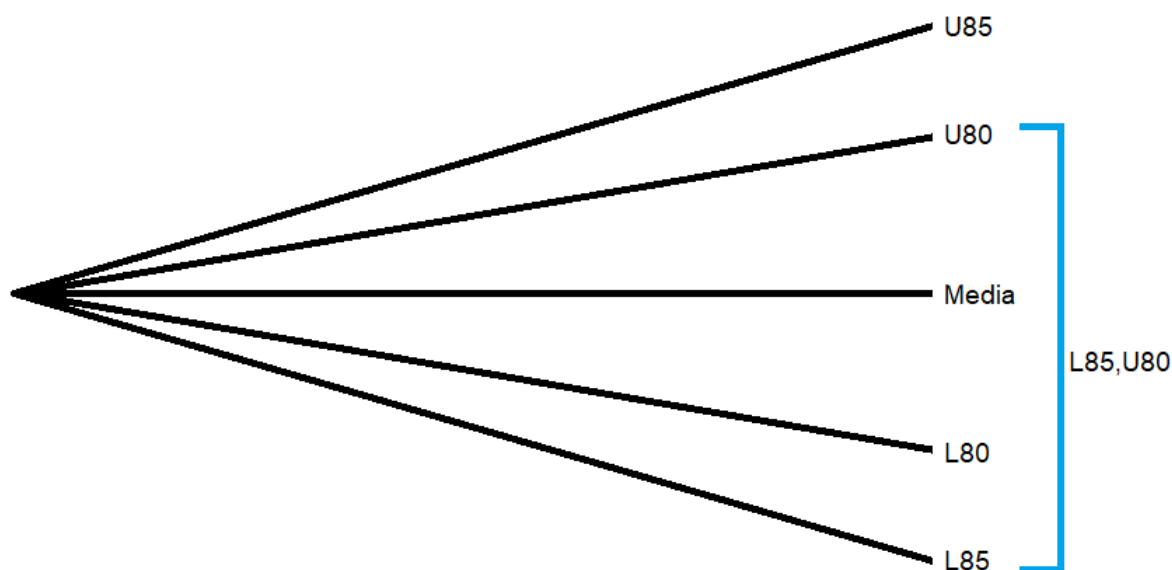


Figura 4.7: Diagrama de los intervalos de confianza

Finalmente con los valores de  $q_1 = 85$  y  $q_2 = 80$  se hizo una prueba aleatoria (eliminando la semilla). Las materias que elegimos para dicha prueba son: *Estadística III*, *Teoría del Seguro*, *Cálculo Diferencial e Integral I*, *Investigación de Operaciones*, *Geometría Moderna I*, *Geometría Analítica II*, *Lógica Matemática I*, *Cálculo Diferencial e Integral III*, *Estadística I*, *Bases de Datos*, *Matemáticas Financieras*, *Cálculo Diferencial e Integral II*, *Probabilidad I*, *Probabilidad II* y *Procesos Estocásticos I*. Los resultados de la prueba aleatoria los podemos ver en la figura 4.8. El promedio de la desviación estándar de todas las materias es 0.48.

Materia	Intervalo	Min	Media	Max	sd	Seg
Estadística III	L85,U80	-2.4053333	-0.062628178	1.0000000	0.7625277	6.68
Teoría del Seguro	L85,U80	-0.7475410	-0.007189294	0.9685714	0.2066836	4.46
Cálculo Diferencial e Integral I	L85,U80	-1.8055556	-0.150509219	0.8544828	0.5743739	7.67
Investigación de Operaciones	L85,U80	-1.5109589	-0.113587042	0.5523364	0.4197767	8.46
Geometría Moderna I	L85,U80	-1.0729730	0.062571864	1.0000000	0.3758780	5.20
Geometría Analítica II	L85,U80	-1.6896552	-0.075029650	1.0000000	0.6112092	7.42
Lógica Matemática I	L85,U80	-1.4857143	0.009737679	1.0000000	0.4441402	6.46
Cálculo Diferencial e Integral III	L85,U80	-1.6142857	-0.118954103	0.8689189	0.5222578	7.26
Estadística I	L85,U80	-1.8022222	-0.045271946	0.9751515	0.5525782	6.57
Bases de Datos	L85,U80	-0.5000000	0.030660747	0.8080000	0.1858422	4.25
Matemáticas Financieras	L85,U80	-0.8403974	0.030471884	0.8584270	0.2544040	4.69
Cálculo Diferencial e Integral II	L85,U80	-3.2192308	-0.053958329	1.0000000	0.6681179	6.77
Probabilidad I	L85,U80	-1.6750000	-0.047856672	0.9188034	0.4289253	5.91
Probabilidad II	L85,U80	-1.9750000	-0.124282718	1.0000000	0.7012144	7.28
Procesos Estocásticos I	L85,U80	-1.8096154	-0.109287138	0.7419643	0.5144173	4.86

Figura 4.8: Matriz con medidas de dispersión de prueba aleatoria

### 4.3. Simulación de la demanda de alumnos

En la figura 4.9 vemos ...

▲	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 4.9: *Matriz con alumnos corregidos*

En la figura 4.10 vemos...

▲	20202
7-8	0
8-9	0
9-10	61
10-11	1
11-12	8
12-13	122
13-14	0
14-15	0
15-16	0
16-17	0
17-18	31
18-19	29
19-20	6
20-21	132
21-22	0

Figura 4.10: *Vector con demanda simulada para el 2020-2*



## 4.4. Simulación de tamaño de grupos

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

## 4.5. Obtención de información para solicitudes

Antes de iniciar las simulaciones de elección de materia y de horario obtuvimos un vector y una matriz con la información de las materias y de los profesores, respectivamente.

En el caso de las materias, el vector *vec\_nom\_materias\_total* lo obtuvimos a partir de la matriz *m\_grande\_total* del semestre 2008-1 al 2020-1. No tiene nombres repetidos. Tiene 333 materias.

En el caso de los profesores, la matriz *mat\_nom\_prof\_total* tiene 2 columnas. En la primer columna se tienen los nombres de todos los profesores que han impartido clase desde el semestre 2015-1 hasta el 2020-1. Dichos nombres los obtuvimos de la matriz *m\_grande\_total* de los semestres correspondientes.

En la segunda columna de la matriz, se tiene un 1 si el profesor es de tiempo completo y un 0 si no. Para llenarla ingresamos a la página <http://www.matematicas.unam.mx/index.php/nosotros/profesores-de-tiempo-completo> del Departamento de Matemáticas. Con la aplicación *SelectorGadget* seleccionamos el vector con el nombre de los profesores de tiempo completo. En la figura 4.11 podemos ver el código CSS que utilizamos para obtener los datos en R. También observamos que se seleccionaron 94 profesores.



Figura 4.11: *Profesores de tiempo completo: SelectorGadget*

Al extraer la información en R obtuvimos un vector con 94 entradas. En la figura 4.12 podemos ver los primeros 20 valores del vector. Notamos que cada entrada del vector inicia con los caracteres `\n\t\t\t\t\t`. Estos caracteres, en la presentación final de la página de internet, indican un salto de línea y las tabulaciones o espacios que se tienen de izquierda a derecha.

```

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

```

Figura 4.12: *Vector de profesores de tiempo completo*

Limpiamos los datos para obtener un vector que sólo tuviera los nombres de los profesores. Eliminamos el título de cada uno porque en los horarios publicados en las páginas de la FC 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 a mano 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 nombre incompleto. En la tabla 4.2 vemos los nombres que aparecen en las páginas de FC comparados con los que aparecen en la página del Departamento de Matemáticas.

Nombre en páginas de FC	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 Lluís Puebla	Emilio Lluís 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: *Diferencias en nombres de profesores de tiempo completo*

#### 4.6. SIMULACIÓN DE SOLICITUDES DE PROFESORES OCULTA Y PSEUDO-REAL<sup>47</sup>

Finalmente en la matriz *mat\_nom\_prof\_total* se tiene la información de 1387 profesores de los cuales 94 son profesores de tiempo completo.

Algunas notas a considerar de esta matriz son:

- Hay profesores que se repiten por diferencia de acentos. Ej. *César Alejandro Arellano Ruíz, Luis Eduardo García Hernández*
- Hay profesores que se repiten por tener a lado el nombre de los ayudantes. Ej. *Fermín Alberto Viniegra Heberlein, Edgar Vázquez Luis*
- Puede haber profesores que ya no impartan clases en la FC.

### 4.6. Simulación de solicitudes de profesores oculta y pseudo-real

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

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

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

#### 4.6.1. Simulación de elección de materia

#### 4.6.2. Simulación de elección de horario

### 4.7. Simulación de esqueletos

Matriz de 2 columnas (Materia-Horario). En el renglón  $i$  se tiene la información de cada grupo simulado para  $t+1$ .

Se utiliza una matriz auxiliar de 3 columnas (Materia-Horario-Demanda\_Alum). En el renglón  $i$  se tiene la información del número de alumnos simulados para la hora y materia correspondientes.

## 4.8. Calificación de esqueletos de horario

$$\begin{aligned}
 L_{materia} &= -1 \text{ por cada materia no impartida} \\
 x &= \text{promedio} \\
 y &= \text{cupo} \\
 L_{dif\_p\_c}(x, y) &= \begin{cases} \frac{a}{190}(x - y) & \text{si } x < y \\ -\frac{b}{190}(x - y) & \text{si } x \geq y \end{cases} \\
 a &= 0.5 \\
 b &= 0.8 \\
 L_{categoria}^1(mat, solicitud) &= -c(categoria - 1) \\
 L_{categoria}^2(mat, solicitud) &= -c_1(categoria)
 \end{aligned}$$

Al momento de simular las solicitudes de materias para los profesores suponemos que la que está en primer lugar es la materia que más quiere dar, en segundo lugar, la segunda que quiere dar.

Primero asignar grupos a los profesores de tiempo completo. Después asignar grupos faltantes a los profesores de asignatura.

$L_{materia}$  es la penalización por no tener en el esqueleto una materia que necesitamos.

$L_{dif\_p\_c}$  es la penalización en la asignación de salones. Se tiene un grupo de tamaño  $x$  y un salón con capacidad  $y$ . Se penaliza con  $\frac{a}{190}$  veces la diferencia entre  $x$  y  $y$  cuando el tamaño del grupo es menor a la capacidad del salón y se penaliza con  $-\frac{b}{190}$  veces la diferencia entre  $x$  y  $y$  cuando el tamaño del grupo es mayor a la capacidad del salón.

El esqueleto depende de la demanda de alumnos y de las solicitudes de los profesores.

Primero se asignan materias a los profesores de tiempo completo y después a los de asignatura.

Los profesores de asignatura pueden quedarse sin materias asignadas.

Penalizaciones:

1. Si algún profesor pidió alguna materia y no se la dieron.
2. Si hay alumnos que necesitan una clase a alguna hora y no existe profesor que la imparta.
3. Con  $\alpha \times num\_alumnos\_faltantes$  por cada alumno que te faltó en cada hora-materia que tenías que dar.  $\alpha > 0$
4. Con  $\beta \times num\_alumnos\_sobrantes$  por cada alumno que te pasaste en cada hora-materia que tenías que dar.  $\beta > 0$

Queremos el esqueleto con el menor valor en  $\alpha + \beta$

Con esto se obtiene un nuevo esqueleto.

## **4.9. Generación de asignaciones**

Matriz de 3 columnas (Materia-Horario-Profesor), la cual tiene la información de las asignaciones. A cada renglón de la matriz de esqueletos se agrega un profesor. Se genera con el esqueleto obtenido del proceso del AG y de las solicitudes de los profesores.

### **4.9.1. Calificación de asignaciones de grupo**



## Capítulo 5

# Teoría del Algoritmo Genético aplicado a los horarios

\*\*\* ESCRIBIR ACERCA DE LA TEORÍA DE AG \*\*\*

El algoritmo genético actualmente se utiliza para resolver problemas de optimización tanto discretos como continuos. Se basa en el mecanismo de la selección natural de Darwin, el cual nos indica que el individuo más apto sobrevive, por lo que entre mejores sean los padres, mejor es la descendencia.

Definimos a un cromosoma como una posible solución al problema. En nuestro caso representamos a un cromosoma por medio de una matriz con  $j\_materias$  renglones y con 3 columnas las cuales representan la asignación de profesor, día y salón, respectivamente, por lo que el renglón  $j$  indica que la materia  $j$  es impartida por el profesor  $i$ , el día  $t$ , en el salón  $k$ .

El valor de adaptabilidad  $fit(x)$ , de cada cromosoma, se asigna al evaluar su utilidad en la función objetivo, entre mejor sea el cromosoma, más alto será su valor de adaptabilidad. Los mejores cromosomas de la población actual pasan directamente a la siguiente generación. Se dice que la población evoluciona por medio de tres operadores hasta una condición de paro, los operadores son: *selección*, *entrecruzamiento* (*crossover*) y *mutación*.

Los pasos del algoritmo se muestran a continuación:

1. Se inicia con un grupo de cromosomas generados aleatoriamente, a los cuales se les calcula su valor de adaptabilidad
2. La probabilidad de que el cromosoma  $k$  sea elegido para el entrecruzamiento (*crossover*), es:

$$p_k = \frac{fit(x)}{\sum_{h=1}^{pop} fit(h)} \quad \text{donde } pop \text{ es el tamaño de la población de cromosomas}$$



3. En el entrecruzamiento se mezclan dos padres para generar nuevas soluciones. Se genera un número aleatorio entre cero y uno,  $r$ , si  $r < 0.6$  la primer columna de  $M_{ij}$  y la primer columna de  $M_{ti}$  del padre 1 se copian en la nueva solución, las demás columnas se llenan con las columnas del padre 2. Si la nueva solución no es factible, en la matriz  $M_{ij}$ , si alguna materia tiene asignada dos profesores, se selecciona uno de ellos de manera aleatoria y el otro se elimina de esa asignación; en caso de que alguna materia no tenga profesor asignado, se le asigna uno aleatoriamente.
4. Se actualiza la matriz  $M_{ti}$ .
5. Se aplica el operador *mutación*, se selecciona un profesor de manera aleatoria y se cambia el día en el que más tiene clase por el día que menos clases imparte. Ésto se aplica para cada profesor de manera aleatoria, sin repetición.
6. Una vez generadas las nuevas soluciones se elige la mejor entre todas ellas.

## 5.1. Ciclo de la evolución natural



Figura 5.1: Algoritmo Genético

### 5.1.1. Selección

### 5.1.2. Cruce

### 5.1.3. Mutación

### 5.1.4. Reemplazamiento

## 5.2. Algoritmo Genético aplicado a la generación de esqueletos de horario

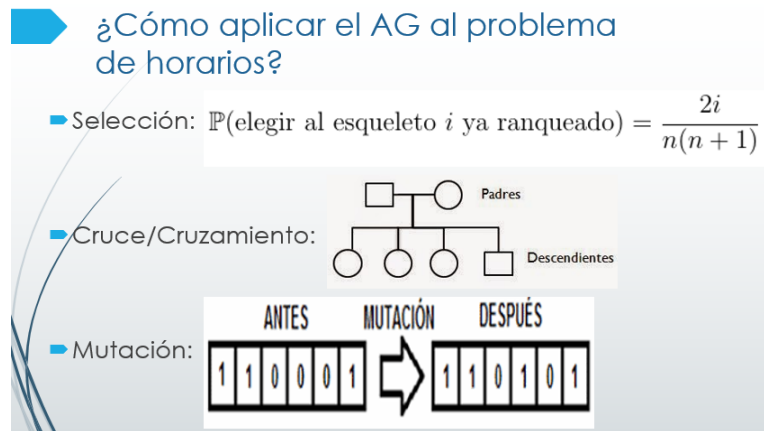


Figura 5.2: Algoritmo Genético aplicado

## 5.3. Algoritmo Genético aplicado a la generación de asignaciones de grupos



## **Capítulo 6**

### **Resultados del Algoritmo Genético**



## **Capítulo 7**

### **Comportamiento de la selección**



# Capítulo 8

## Conclusiones

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

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

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

Con el fin de encontrar más posibles aplicaciones del programa realizado en este trabajo, se buscaron diferentes páginas de horarios en distintas facultades de la UNAM y de otras universidades. No se pudieron encontrar páginas con todas las características que tienen las páginas de la FC. Algunas de las páginas que se encontraron son las siguientes:

- *Facultad de Filosofía y Letras UNAM*: Se encontró una estructura en las páginas web con las cuales se puede acceder a la información por carrera, pero no se puede acceder a la información de semestres anteriores y dentro de éstas páginas no se pueden encontrar el número de alumnos inscritos por cada materia, por lo que no sería posible una simulación del número de alumnos.

<https://servicios-galileo.filos.unam.mx/horarios/ordinarios/1354>

<https://servicios-galileo.filos.unam.mx/horarios/ordinarios/1355>

<https://servicios-galileo.filos.unam.mx/horarios/ordinarios/1359>

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

<https://www.ssa.ingenieria.unam.mx/horarios.html>

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

- *FES Acatlán (Actuaría)*: En la siguiente página web se pueden descargar los horarios del semestre en curso en un archivo de Excel el cual no contiene información del número de alumnos inscritos en el grupo, tampoco se puede obtener información de



semestres anteriores. No se puede utilizar la aplicación *SelectorGadget* para obtener la información.

<http://www.actuaria.acatlan.unam.mx/>

- *FES Iztacala (Psicología)*: Se encontró una estructura en las páginas web con las cuales se puede acceder a la información en archivos pdf de algunos semestres, dependiendo si el semestre en curso es par o impar. No se puede utilizar la aplicación *SelectorGadget* para obtener la información.

[https://psicologia.iztacala.unam.mx/avisos2020/horarios21\\_1/21-1\\_3-TERCER%20SEMESTRE.pdf](https://psicologia.iztacala.unam.mx/avisos2020/horarios21_1/21-1_3-TERCER%20SEMESTRE.pdf)

[https://psicologia.iztacala.unam.mx/avisos2020/horarios21\\_1/21-1\\_5-QUINTO%20SEMESTREv1108.pdf](https://psicologia.iztacala.unam.mx/avisos2020/horarios21_1/21-1_5-QUINTO%20SEMESTREv1108.pdf)

- *Centro de Nanociencias y Nanotecnología (Nanotecnología)*: Al igual que en el caso anterior la información de las páginas que se muestran a continuación están en archivos pdf por lo que no se puede utilizar la aplicación *SelectorGadget* para obtener la información.

<https://nanolic.cnyn.unam.mx/sitio/wp-content/uploads/2020/09/H-1A-2021-1.pdf>

<https://nanolic.cnyn.unam.mx/sitio/wp-content/uploads/2020/09/H-1B-2021-1.pdf>

<https://nanolic.cnyn.unam.mx/sitio/wp-content/uploads/2020/09/H-3A-2021-1.pdf>

- *Facultad de Química UNAM*: En la siguiente página web se pueden seleccionar todas las materias impartidas en la facultad o por carrera.

[http://escolares.quimica.unam.mx/Horarios/hor\\_def\\_e2.php4](http://escolares.quimica.unam.mx/Horarios/hor_def_e2.php4)

Una vez que se eligió alguna opción, se muestra un listado, en la siguiente url, con las posibles materias que se pueden elegir.

[http://escolares.quimica.unam.mx/Horarios/hor\\_def\\_pre\\_e2.php4](http://escolares.quimica.unam.mx/Horarios/hor_def_pre_e2.php4)

Finalmente se accede a la información con la siguiente página web.

[http://escolares.quimica.unam.mx/Horarios/hor\\_tot\\_e2.php4](http://escolares.quimica.unam.mx/Horarios/hor_tot_e2.php4)

No importa las opciones que se elijan, siempre se obtienen esas mismas urls por lo que no hay alguna estructura para poder buscar la información automáticamente.

- *ESIME Zacatenco IPN (Ingeniería en Control y Automatización)*: Se encontró una estructura en las páginas web pero no se puede encontrar el número de alumnos inscritos por materia por lo que no es posible realizar una simulación del número de alumnos.

<http://horarios.esimez.ipn.mx/horarios/VHorGpoAl.aspx?Gpo=1AM8&PaId=57>

<http://horarios.esimez.ipn.mx/horarios/VHorGpoAl.aspx?Gpo=1AV1&PaId=57>

- *ITAM*: En este caso se debe de seleccionar una materia y luego se despliega la información, sin importar la selección de la materia, las url son las mismas por lo que no se tiene una estructura en las páginas web.

[http://escolar1.rhon.itam.mx/licenciaturas/horarios/seleccion\\_03.asp](http://escolar1.rhon.itam.mx/licenciaturas/horarios/seleccion_03.asp)

[http://escolar1.rhon.itam.mx/licenciaturas/horarios/pormateria\\_03.asp](http://escolar1.rhon.itam.mx/licenciaturas/horarios/pormateria_03.asp)



09:14:54 p. m. del 25-septiembre-2020

Los grupos programados para el semestre OTOÑO 2016 LICENCIATURA de la materia CALCULO DE PROBABILIDADES, I son:

DEPTO.	CLAVE	GRUPO	TEORÍA O LABORATORIO	NOMBRE	PROF.	CRÉDITOS	HORARIO	DÍAS	SALÓN	CAMPUS	COMENTARIOS
EST	14101	001	T	CALCULO DE PROBABILIDADES, I	VICTOR MANUEL ARMANDO AGUIRRE TORRES	6	10:00-11:30	LU MI	RH314	RIO HONDO	
EST	14101	002	T	CALCULO DE PROBABILIDADES, I	ANA MEDA GUARDIOLA	6	10:30-12:00	MA JU	RH313	RIO HONDO	
EST	14101	003	T	CALCULO DE PROBABILIDADES, I	ERICK MIER MORENO	6	08:30-10:00	LU VI	RHPB2	RIO HONDO	
EST	14101	004	T	CALCULO DE PROBABILIDADES, I	LUIS ENRIQUE NIETO BARAJAS	6	13:00-14:30	LU MI	RH314	RIO HONDO	
EST	14101	005	T	CALCULO DE PROBABILIDADES, I	MIGUEL ANGEL MENDEZ ANTONIO	6	08:30-10:00	LU MI	RH313	RIO HONDO	

Figura 8.1: *ITAM Probabilidad I*

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

<https://cienciasquimicas.lasalle.mx/wp-content/uploads/2020/08/QFB-291.pdf>

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

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

- *Universidad Panamericana*: No se encontraron horarios de materias, sólo de exámenes y de entrenamientos.

[https://www.up.edu.mx/sites/default/files/fechas\\_de\\_examen\\_humanidades\\_1202.pdf](https://www.up.edu.mx/sites/default/files/fechas_de_examen_humanidades_1202.pdf)

<https://www.up.edu.mx/en/media/22960>

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

- *Biología*:

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

- *Ciencias de la Tierra:*

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

- *Física:*

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

- *Física Biomédica:*

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

- *Manejo Sustentable de Zonas Costeras:*

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

# Apéndices

## 8.1. Observaciones / Notas

1. La matriz `mat_posibles_url` se define con un tamaño fijo antes de correr el algoritmo para que no se demore por tener un objeto que va cambiando de tamaño, por lo que al final de haberle aplicado la función se le deben de quitar los renglones que no tienen información.
2. La función `casos_alumnos` convierte los *NA* de la columna *Alumnos* de `m_grande` en ceros pero al generar `m_grande_total` y pasarla por la función `limpia_m_grande` se eliminan los *NA* y se cambian por ceros por lo tanto no es necesaria la función `casos_alumnos`, basta pasar la columna correspondiente a *Alumnos* de `m_grande_total`.
3. Cuando se hacen comparaciones se toman los valores reales y se les restan los valores simulados ( $Reales - \mathbb{E}[Simulados]$ )
4. Con las gráficas *heatmap* se revisa si el modelo es adecuado o si se debe modificar algo. Se espera que las gráficas sean de color claro ya que nos interesa que el número de grupos y alumnos simulados se parezca al real.
5. Se tienen dos tipos de matrices las cuales llamaremos *m\_objetivo* y *m\_definición*; las matrices *m\_objetivo* son las que tienen la información que se utiliza para la asignación; las matrices *m\_definición* nos sirven para dos cosas:
  - a) Respaldo de la descripción de cada columna
  - b) Para guardar los índices en los que se encuentran las columnas
6. Las matrices tipo *m\_definición* son:
  - a) `mat_def_columnas_MG`
  - b) `mat_def_grupos_reales`
  - c) `mat_def_grupos_simulados`
7. Las matrices tipo *m\_objetivo* son:
  - a) `m_grande`
  - b) `m_grande_total`
  - c) ...

8. La función *checa\_ind\_materia* se encarga de obtener 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:

```

1  for(j){
2    for(i){
3      m[i,j]
4    }
5  }
6
```

Código 8.1: 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 8.2: Resumen de clases de ingles antes de modificación

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

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

```

1  gen_asignacion_completa <- function(sem_ini, sem_fin){
2    # Se carga y se limpia la lista de urls (para no tener
    paginas sin informacion,...)
3    list_url <- Actualiza_list_url(list_url)
4
5    # Se obtiene "m_grande" y se genera un archivo para cada
    semestre
6    for(k in 1:length(semestres)){
7      sem_info <- semestres[k]
8      directorio_info[k] <- gen_m_grande(sem_info, list_url)
9    }
10
11   # Se genera el esqueleto del semestre que se quiere obtener
12   mat_esqueleto <- gen_esqueleto(directorio_info, param)
13
14   # Se genera la matriz de solicitudes de todos los profesores
15   mat_solicitudes <- gen_solicitudes(param)
16
17   # Se genera la matriz de asignaciones de todos los
    profesores
18   mat_asignaciones <- gen_asignacion(mat_esqueleto, mat_
    solicitudes, param)
19
20   return(mat_asignaciones)

```

21 }  
 22 }  
 23 }

### Código 8.2: 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 FC 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.

Horarios 2021-1

Navegación: Ciencias » Horarios 2021-1 » Actuaría » Modelos no Paramétricos y de Regresión.  
Anterior: Métodos Cuantitativos en Finanzas Siguiente: Análisis del México Contemporáneo

Actuaría (plan 2015)

Modelos no Paramétricos y de Regresión, Sexto Semestre

Buscar en grupos

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 Dionei Alonso Rosas Sanchez ma ju 17 a 18

Presentación

Grupo 9226

Profesor Francisco Sánchez Villarreal lu mi vi 18 a 19

Ayudante Santiago Lara Jiménez ma ju 18 a 19

Presentación

Figura 8.3: Ejemplo de horarios de semestre 2021-1

23. Notas de T26

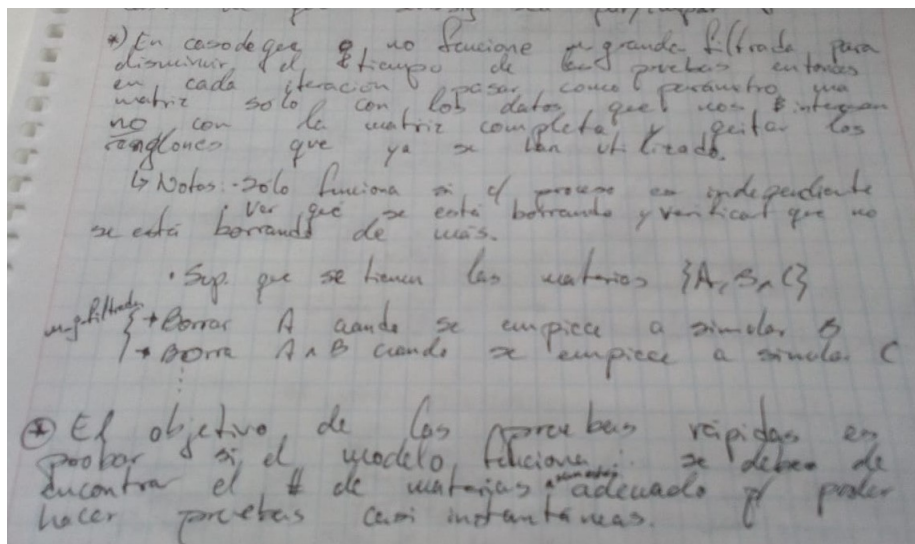


Figura 8.4: Notas de T26

24. En caso de tener subsecciones: entre 3 y 4
25. 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.
26. Sea  $D = \frac{r-s}{s}$ , donde  $r$  son datos reales,  $s$  datos simulados y  $D$  la diferencia relativa, se busca que  $D \in \left[-\frac{1}{2}, \frac{1}{2}\right]$ .
27. 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 8.5: Ejemplo de Roxygen

28. 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.
29. Escribir párrafos de 2 a 3 enunciados completos, no dejar enunciados solos a menos que contengan información muy importante.
30. En caso de tener más de 10 referencias bibliográficas utilizar *Mendeley* para genera un archivo *.bib* y ponerlo en la tesis para tener la bibliografía.
31. Cuidar el tamaño de letra en las gráficas que se pongan
32. No poner abreviaturas en los títulos.
33. Ser muy directa al escribir, pero explicar mucho más (platicar más). No hacer enunciados tan largos. Escribir una idea por enunciado. No sólo escribir en párrafos, utilizar listas, tablas, ...
34. La imagen 3.4 tiene título en inglés, se tienen 2 opciones: dejarlo así o buscar cómo cambiarlo.
35. Recordar la diferencia entre:
  - Número de alumnos inscritos
  - Número de alumnos reales



- Número de alumnos que toman clase por cada horario (no se toman en cuenta los alumnos que empalman clases)
36. Para la elección de  $q_1$  y  $q_2$  se debe darle prioridad a la varianza no al mín y al máx porque se pueden tener casos en los que el mín y el máx estén muy cercanos a cero (gráfica superior) pero su varianza es grande. Queremos que la varianza se encuentre alrededor del cero (gráfica inferior).

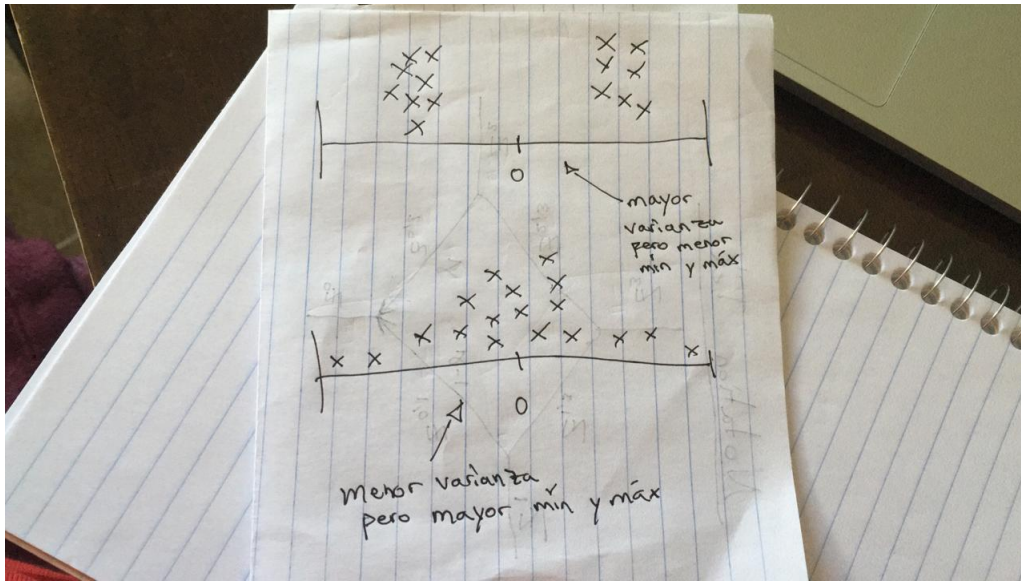


Figura 8.6: Ejemplo de varianza

37. Preferir sacrificar el B/N en las imágenes impresas para tener una mejor versión digital a color.
38. Guardar figuras hechas en R con el comando: `dev.print(pdf, "Figures/Fig_Examples_of_GB_distributions", height=5)`
39. Arrigo dijo que posiblemente alguien se va a quejar de no tomar en cuenta la preferencia de los profesores al realizar las solicitudes.
40. 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.
41. Cláusula 99 CCTPA: Ayuda para la impresión de la tesis.

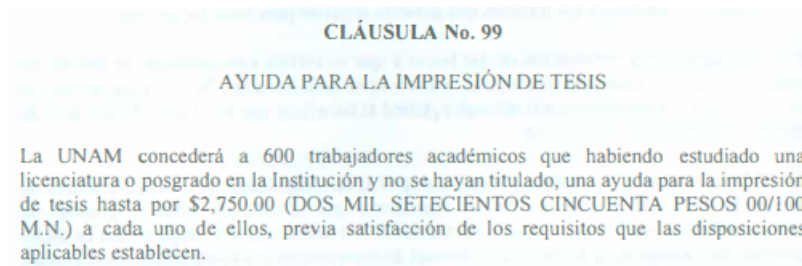


Figura 8.7: *Cláusula 99 CCTPA: Ayuda para la impresión de la tesis*

42. 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

43. 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 (9).

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

45. 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  $T_{32xx}$

46.

47.

48.

49.

50.

## 8.2. Resultados útiles

**Definición 8.1.** *Estimador máximo verosímil de  $\lambda$*

Sean  $X_1, X_2, \dots, X_n$  una muestra aleatoria de una población con función de densidad de probabilidad  $\text{Poisson}(\lambda)$ . Su función de densidad es:

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad (8.1)$$

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

*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  $\lambda$*

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

*Igualamos a cero*

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

*Despejamos  $\lambda$*

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

*Derivamos otra vez*

$$\frac{\partial^2}{\partial \lambda^2} \ln \mathcal{L}(\underline{X}; \lambda) = -\frac{\sum_{i=1}^n x_i}{\lambda^2} < 0$$

$$\therefore \hat{\lambda} = \bar{x} \text{ es el estimador máximo verosímil}$$

### 8.3. Abreviaturas

ABREVIATURA	SIGNIFICADO
CdC	Ciencias de la Computación
ESIME	Escuela Superior de Ingeniería Mecánica y Eléctrica
FC	Facultad de Ciencias de la UNAM
FES	Facultad de Estudios Superiores
ITAM	Instituto Tecnológico Autónomo de México
MatAp	Matemáticas Aplicadas
UNAM	Universidad Nacional Autónoma de México
URL	Uniform Resource Locator
a	b

Tabla 8.1: Abreviaturas

# Bibliografía

- [1] Casella G., (2006), *Statistical Inference*, Thomson Press
- [2] Chatfield C. y Xing H., (2019), *The Analysis of Time Series An Introduction with R*, Chapman & Hall/CRC
- [3] Gibbons J. D. y Chakraborti S., (2011), *Nonparametric Statistical Inference*, Chapman & Hall/CRC
- [4] Montgomery D., Jennings C. y Kulahci M., (2015), *Introduction to Time Series Analysis and Forecasting*, Wiley
- [5] Rincón L., (2007), *Curso intermedio de probabilidad*, UNAM
- [6] Rubinstein R. y Kroese D., (2016), *Simulation and the Monte Carlo Method*, Wiley
- [7] Shumway R. y Stoffer D., (2017), *Time Series Analysis and Its Applications: With R Examples*, Springer
- [8] Yazdani M., Naeri B. y Zeinali E., (2017), *Algorithms for university course scheduling problems*, Tehnički vjesnik 24, Supl. 2, 241-247
- [9] <https://estadisticafciencias.files.wordpress.com/2019/08/introduccion-a-la-estadistica-versi3n-final.pdf>