



Universidad Nacional Autónoma de México

Maestría en Ciencias Matemáticas

Modelos Lineales Generalizados

Proyecto

Profesora Dra. Silvia Ruiz-Velazco Acosta

Elaborado por:

Alarcón González Edgar Gerardo

Índice

1. Introducción

2. Descripción y declaración de las variables

3. Análisis Descriptivo de los Datos

3.1 Por variable

3.2 Por conjunto de variables

4. Modelos

4.1 Para los Egresados

4.1.1 Regresión Logística / Logit

4.1.2 Regresión Probit

4.2 Para los Graduados

4.2.1 Regresión Logística / Logit

4.2.2 Regresión Probit

5. Selección de Modelo

5.1 Comparativas de Modelo

5.1.1 Para los egresos

5.1.2 Para los graduados

5.2 Interpretación del modelo final

5.2.1 Para los egresados

5.2.2 Para los graduados

6. Predicciones

7. Conclusiones

1. Introducción

Los datos que se trabajan en este documento son datos de alumnos/aspirantes que entraron a la “Maestría en Ciencias Matemáticas” por parte del programa “Maestría y Doctorado en Ciencias Matemáticas y de la Especialización en Estadística Aplicada” de la Unidad de Posgrado de la Universidad Nacional Autónoma de México (UNAM). Como parte de la labor administrativa, es de interés que los alumnos/aspirantes logren el objetivo de titularse, esto debido al alto nivel académico que se maneja y reporta a instituciones como el Consejo Nacional de Ciencia y Tecnología (CONACyT).

Por tal motivo, la elección correcta de los aspirantes resulta ser un aspecto fundamental para progresar con dicho rendimiento y desempeño académico. De tal manera que se ha recopilado información estadística de algunos de los aspirantes a entrar a la maestría de los semestres 2015-1 al 2019-2 con el **objetivo** de buscar alguna relación que favorezca la posibilidad de graduarse del grado académico al que se aspira. Todo esto se realizará utilizando los conceptos más fundamentales de Modelos Lineales Generalizados (MLG). El software que estaremos trabajando será R.

2. Descripción y declaración de las variables

Teniendo bajo contexto que los datos vienen de 301 alumnos/aspirantes a la maestría en ciencias matemáticas. Debido al formato de fechas y la manera en que los datos fueron entregados, se decidió hacer un pequeño pre-procesamiento de los mismos en Excel.

```
datos <- read.csv("edgar.csv",na.strings = "",encoding = "UTF-8")
```

Aunque existen individuos con datos faltantes, las mediciones realizadas fueron las siguientes¹.

1. **Semestre:** Semestre en que ingresó a la maestría. Esta variable se expresa en forma de fecha dentro de los datos originales, sin embargo, nosotros la vamos a considerar como variable categórica y será decodificada de tal manera que esta se pueda leer como el semestre escolar. Por ejemplo, el dato 04/ago/2014 será codificado como 2015-1. Hay un total de 10 semestres diferentes.

```
# Semestre
datos$Semestre <- datos$Semestre %>% factor
levels(datos$Semestre) <- c("2019-2", "2016-1", "2016-2", "2015-2", "2018-2",
                           "2017-2", "2015-1", "2019-1", "2018-1", "2017-1")
```

2. **Sexo:** Sexo/Género (biológico) del individuo. En los datos originales, esta variable se encontraba como “M” y “F”, serán codificadas como “Hombre” y “Mujer” respectivamente.

```
# Sexo
datos$Sexo <- datos$Sexo %>% factor
levels(datos$Sexo) <- c("Mujer", "Hombre")
```

3. **Fecha_Examen:** Fecha en que presentó su examen de titulación de la maestría. Esta variable es interesante, será considerada como lo que es, una variable de tipo Fecha que en el caso de R en el fondo la piensa como si fuera del tipo numérico. Esto lo hacemos así por la gran cantidad de fechas diferentes entre individuos. **Más importante aún** es que con esta variable obtendremos otra llamada Graduado, que es con la que veremos si el alumno llegó a la meta final de titularse de la maestría.

```
# Fecha_Examen
datos$Fecha_Examen <- datos$Fecha_Examen %>% as.Date(format="%d/%m/%Y")
```

4. **Posponer:** Ha ingresado a la maestría pero se decide posponer la inscripción este número de semestres. Aunque podríamos pensarla como variable categórica, se ha decidido mantenerla como numérica, al existir de alguna manera una medida de orden en esta variable, en lo personal, hace sentido pensar un coeficiente en un modelo lineal generalizado que apoye la magnitud de la naturaleza de la variable.
5. **Edad:** Edad de ingreso a la maestría. Se tomará del tipo numérico de igual manera por la gran diversidad de edades que hay y el sentido de interpretabilidad que se le puede dar a su coeficiente en un modelo lineal generalizado.

6. Carrera: Licenciatura que realizó. Esta variable es considerada como categórica, sin embargo, observamos la existencia de una gran diversidad carreras de las cuales provienen los aspirantes (Figura 1).

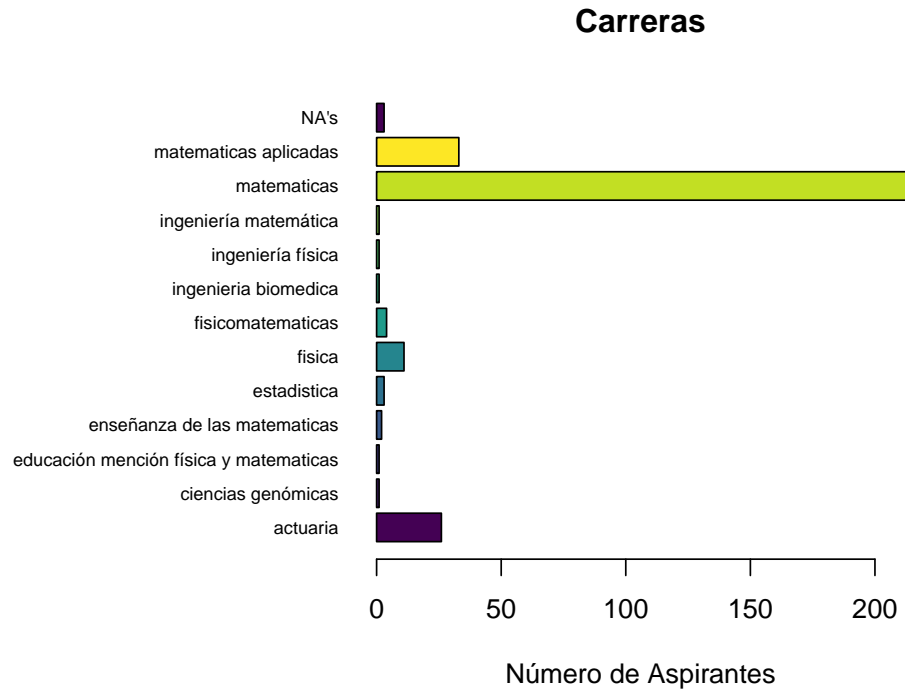


Figura 1: Carreras origen de los aspirantes.

Debido a esto, se decidió re-categorizar la variable, de tal manera que estaremos haciendo implícitamente un suavizamiento sobre la misma. Este procedimiento consistirá en que las carreras que tengan un bajo número de personas entrarán a la categoría “Otras”, mientras que las demás conservarán su nombre original. Al final, terminamos con 5 categorías diferentes Actuaría, Física, Matemáticas, Matemáticas Aplicadas y Otras.

7. Universidad: Universidad de la que proviene. Esta variable al igual que la anterior, presenta una gran diversidad y también es categórica (Figura 2).

Sin embargo, observamos que existe una “Universidad” que tiene muchos más aspirantes que el resto (FC UNAM). De tal manera que el suavizamiento que haremos sobre esta variable será que pertenezcan a esta “Universidad” o pertenezcan a alguna otra.

```
datos$Universidad <- datos$Universidad %>% factor()
levels(datos$Universidad) <- ifelse(datos$Universidad=="FC UNAM", "FC UNAM", "Otra")
```

8. Promedio: Promedio escolar como estudiante de la Licenciatura. Esta variable se considerará del tipo numérico.
9. Tiempo: Tiempo en que realizó la licenciatura. Esta variable se considerará del tipo numérico, similar a la variable Posponer podría pensarse como categoría, pero la dejaremos de tipo numérico con la finalidad de dar interpretación con base en magnitud a su coeficiente en el MLG.
10. Álgebra: Calificación en el examen de álgebra (lineal) para el ingreso a la maestría. Esta variable se considerará del tipo numérico.
11. Cálculo: Calificación en el examen de cálculo para el ingreso a la maestría. Esta variable se considerará del tipo numérico.

¹Se muestra el código de las variables que requirieron codificación para R. Las que no presentan código es porque los datos fueron leídos de la forma deseada para su posterior uso.



Figura 2: Universidades origen de los aspirantes.

12. **Veces:** Número de veces que presentó el examen de admisión. Esta variable se considerará del tipo numérico, similar a la variable *Posponer* podría pensarse como categoría, pero la dejaremos de tipo numérico con la finalidad de dar interpretación con base en magnitud a su coeficiente en el MLG.
13. **Egresado (Respuesta 1):** Terminó el 100 % de los créditos de maestría. Esta es una variable de respuesta binaria, algo que ya nos da una idea de qué tipo de modelos podemos aplicar. Originalmente, esta variable viene expresada como 1's y 0's, lo cuál será codificado a "Sí"s y "No"s respectivamente.

```
datos$Egresado <- datos$Egresado %>% factor()
levels(datos$Egresado) <- c("No", "Sí")
```

14. **Graduado (Respuesta 2):** Se logró (o no) el objetivo final de graduarse de la maestría. Al igual que la otra variable respuesta, esta es del tipo binario y será construida a partir de la variable *Fecha_Examen*, la forma en cómo será construida será tomando como graduado a aquellos que tengan diferente de vacío la variable en cuestión y como no graduado en otro caso. Esto se resumirá a un "Sí" o "No" respectivamente.

```
datos$Graduado <- ifelse(is.na(datos$Fecha_Examen), "No", "Sí") %>% factor()
```

Apoyándonos de estas variables principalmente comenzamos realizando un análisis estadístico de las mismas.

3. Análisis Descriptivo de los Datos

Antes de intentar ajustar algún modelo en particular, vamos a comenzar realizando un breve análisis descriptivo de los datos, esta sección la dividiremos en dos partes importantes. En la primera parte analizaremos algunas de las variables de forma individual con el objetivo de entender su comportamiento y posteriormente en la segunda veremos el comportamiento en conjunto de algunas de las variables con el objetivo de ver si hay autocorrelación o existe alguna posible relación con la respuesta o alguna otra variable categórica.

3.1 Por variable

En particular haremos gráficas de frecuencia de las variables categóricas.

- **Sexo** (Figura 3). En este gráfico, podemos observar que existe un sesgo importante en favor del número de hombres que entran a la maestría. Hay una cantidad considerablemente mayor de este Sexo siendo entre 4 ó 5 veces mayor que la cantidad de mujeres.

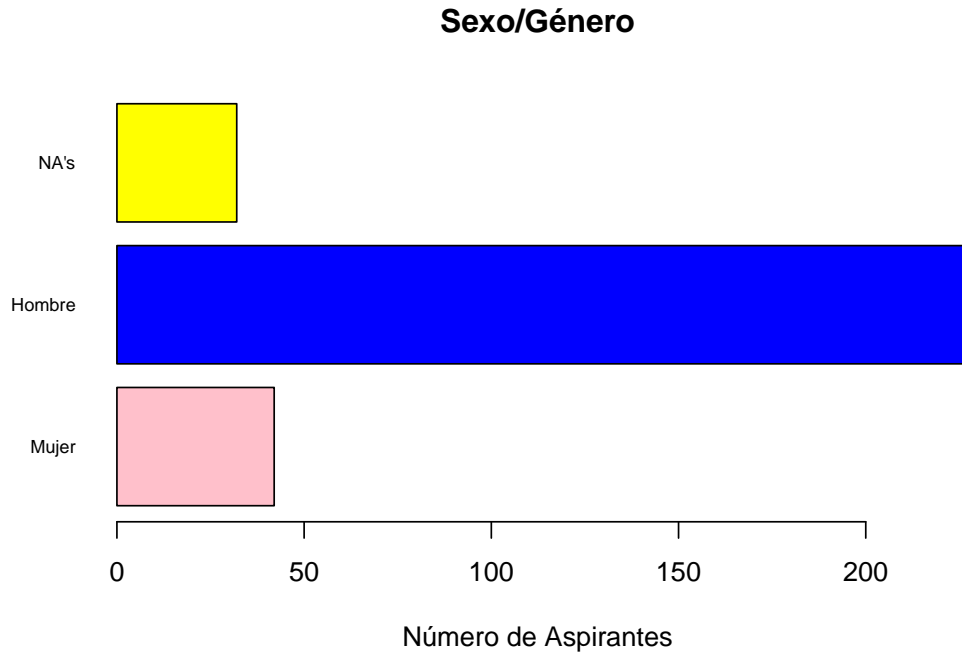


Figura 3: Gráfico de barras por sexo de los aspirantes.

- **Universidad** (Figura 4). Una vez habiendo agrupado, vemos que el sesgo de la cantidad de alumnos que ingresa a la maestría por parte de la facultad de ciencias continúa siendo importante, pero al menos logramos que sea comparable y que la variable fuese mucho más fácil de interpretar y manipular.
- **Egresado** (Figura 5). Esta variable es importante pues de hecho será una de las respuestas que nosotros consideraremos. Vemos que de nuevo existe un sesgo muy importante con los alumnos que en efecto son egresados de la maestría. Esto quiere decir que en la realidad, al menos con los datos que estamos trabajando, hay una probabilidad muy alta de que un alumno, una vez dentro de la maestría, en efecto logre ser egresado. En la **Tabla 1** podemos la cantidad exacta de datos que tenemos. Esto último es importante pues podemos notar que nuestros datos no están balanceados, lo cual puede causar ineficiencias al modelo.

Tabla 1: Frecuencia de 'Egresados'

	Frecuencia	Porcentaje
No	15	0.050
Sí	230	0.764
NA's	56	0.186

- **Graduado** (Figura 6). Esta variable es importante pues de hecho será una de las respuestas que nosotros consideraremos. El comportamiento de esta variable es muy interesante ya que, a diferencia de los egresados, aquí sí hay un balance entre graduados y no graduados. Esto quiere decir que en la realidad, al menos con los datos que estamos trabajando, es casi **equiprobable** que, una vez dentro de la maestría, en efecto logre ser graduado. En la

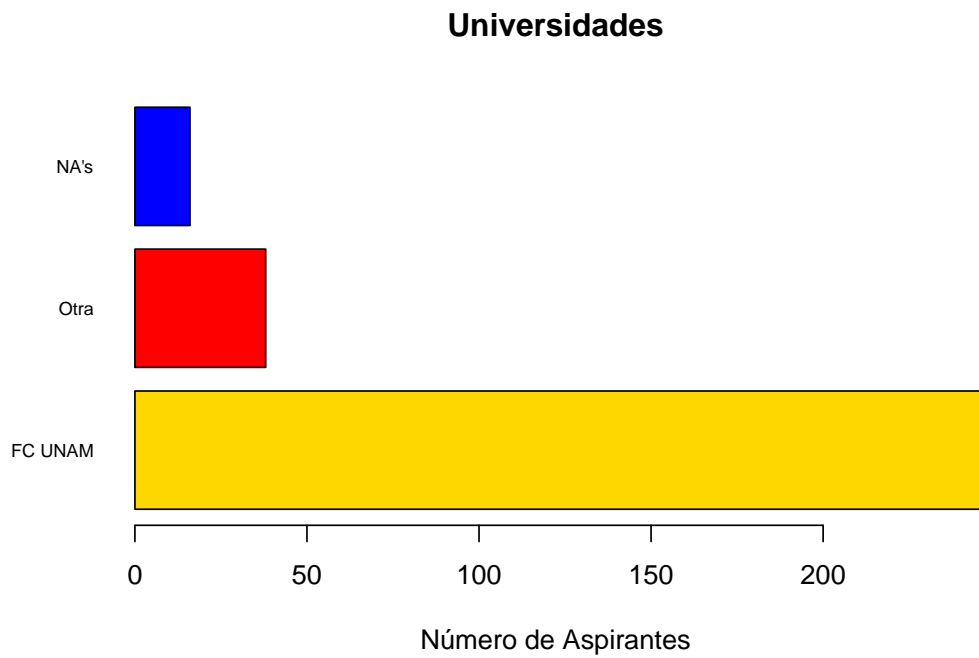


Figura 4: Gráfico de barras de la universidad* de los aspirantes.

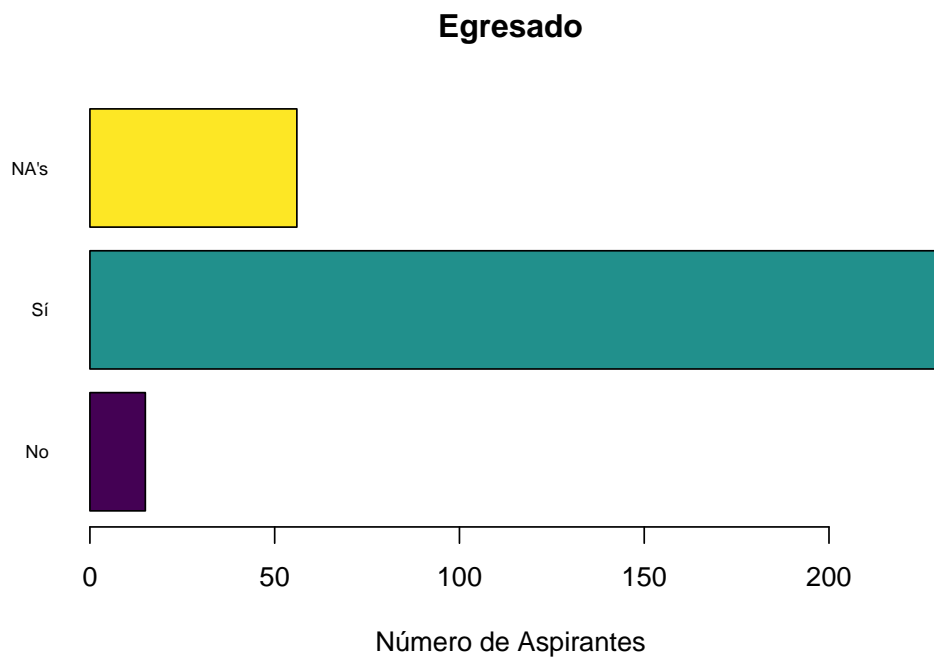


Figura 5: Gráfico de barras de los aspirantes Egresados.

Tabla 2 podemos la cantidad exacta de datos que tenemos. Esto último es importante pues podemos notar que nuestros datos están balanceados, con una ligera ventaja para los que sí se graduaron.

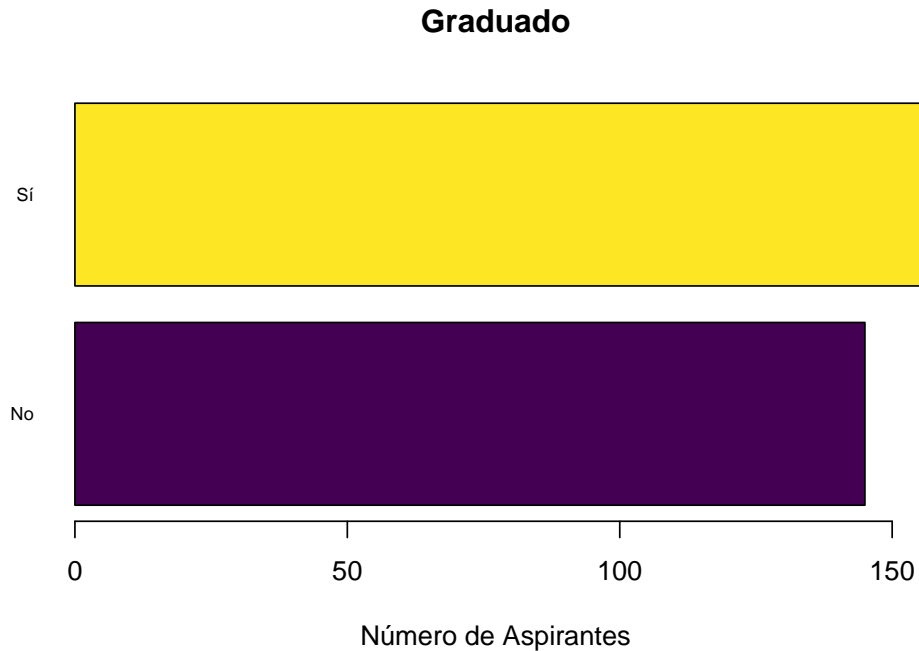


Figura 6: Gráfico de barras de los aspirantes Graduados.

Tabla 2: Frecuencia de 'Graduados'

	Frecuencia	Porcentaje
No	145	0.482
Sí	156	0.518

3.2 Por conjunto de variables

Para esta sección sí tenemos una diversidad de gráficos interesante, incluso incluiremos algunos gráficos marginales que se pudieron incluir en la primera sección, pero para no ser redundantes con eso las tenemos en este apartado.

- **Resultados de examen de cálculo y álgebra (lineal):** Uno de los requisitos más importantes para poder ingresar a la maestría es obtener una calificación satisfactoria en los exámenes de admisión los cuales consisten en las dos áreas de las matemáticas ya mencionadas y que son de las más básicas e importantes en esta rama del conocimiento. Debido a su importancia, vamos a dedicar esta sección a ver su comportamiento contra algunas de las variables categóricas con las que contamos.
 - **Sexo** (Figura 7). En este gráfico, vemos que, parece haber una sutil relación entre la calificación obtenida en el examen y el sexo del aspirante. En particular, parece que un aspirante hombre suele obtener mejores resultados que una mujer. Esta idea recordemos que es condicionada a los datos que tenemos y más aún, que tenemos una muestra muy sesgada en la cantidad de hombres y mujeres.
 - **Universidad** (Figura 8). Este gráfico es muy interesante. Lo que nos muestra es que las calificaciones obtenidas en los exámenes de ingreso parecen tener un comportamiento muy similar entre la Facultad de Ciencias (UNAM) y las demás universidades de origen. Sin embargo, aunque muy leve, se ve que los alumnos de la Facultad de Ciencias tienden a tener una ligera superioridad en los exámenes de álgebra y una casi indistinguible en los de

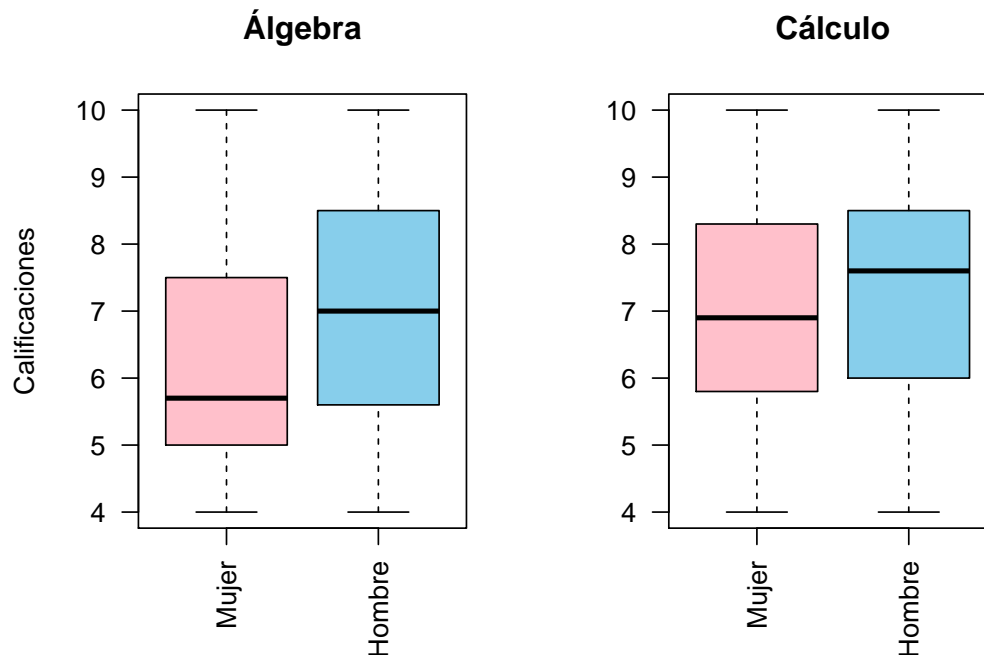


Figura 7: Boxplot de calificaciones de examen por sexo de los aspirantes.

cálculo. Nuevamente, debemos contemplar que hay muchos más alumnos provenientes de la Facultad de Ciencias, entonces, este boxplot puede llegar a representarlos correctamente pero no del todo a las demás universidades.

- **Egresado** (Figura 9). En este gráfico la diferencia es más evidente. Parece ser que una buena forma de explicar si un alumno egresará o no de la maestría, es con base en la calificación obtenida en los exámenes de ingreso. Esta diferencia se ve muy clara en el examen de álgebra (lineal), donde vemos que poco más de la mitad de los alumnos que sí egresaron, obtuvieron una calificación mayor o igual a 7 en este examen, mientras que para los que no egresaron, poco más de 6 de calificación, y principalmente se nota que cuando un alumno obtiene una calificación muy grave en este examen es bastante probable que no egrese de la maestría. Por otro lado, el examen de cálculo es un poco más indistinguible aunque aún se ve muy fuerte la diferencia, parece ser que calificaciones menores a 6.5 son un fuerte indicio de que el aspirante no logrará egresar de la maestría.

Debido a lo visto y mencionado en la Figura 9, continuamos realizando un promedio entre las calificaciones de los exámenes de ingreso y comparándolo también con el promedio obtenido en la carrera.

```
datos <- datos %>% mutate(Promedio_Ex=(Álgebra+Cálculo)/2)
```

Ahora, vamos a ver un gráfico por estatus de egreso de los alumnos en la Figura 10. Aquí, observamos que en efecto la calificación de los alumnos en los exámenes continúa dando una forma de contrastar si un alumno egresará o no de la maestría. Pero algo aún más importante es que esto se ve reflejado de forma aún más evidente por el promedio obtenido en la carrera.

Todo esto nos da una conclusión muy fuerte y es que en efecto, realizar las evaluaciones para evaluar si un aspirante logrará ser egresado es relevante, pero algo igual o quizás más importante es el promedio obtenido en la carrera, herramienta que nos dará buenos indicios de si el aspirante logrará egresar

- **Graduado** (Figura 11). En este gráfico la diferencia es más evidente aún que con el de los egresados. Parece ser que una buena forma de explicar si un alumno egresará o no de la maestría, es con base en la calificación obtenida en los exámenes de ingreso. Esta diferencia se ve muy clara en el examen de álgebra (lineal), donde vemos que poco más de la mitad de los alumnos que sí se graduaron, obtuvieron una calificación mayor o igual a 7 en este examen, mientras que para los que no se graduaron, poco más de 6 de calificación, y principalmente se nota que

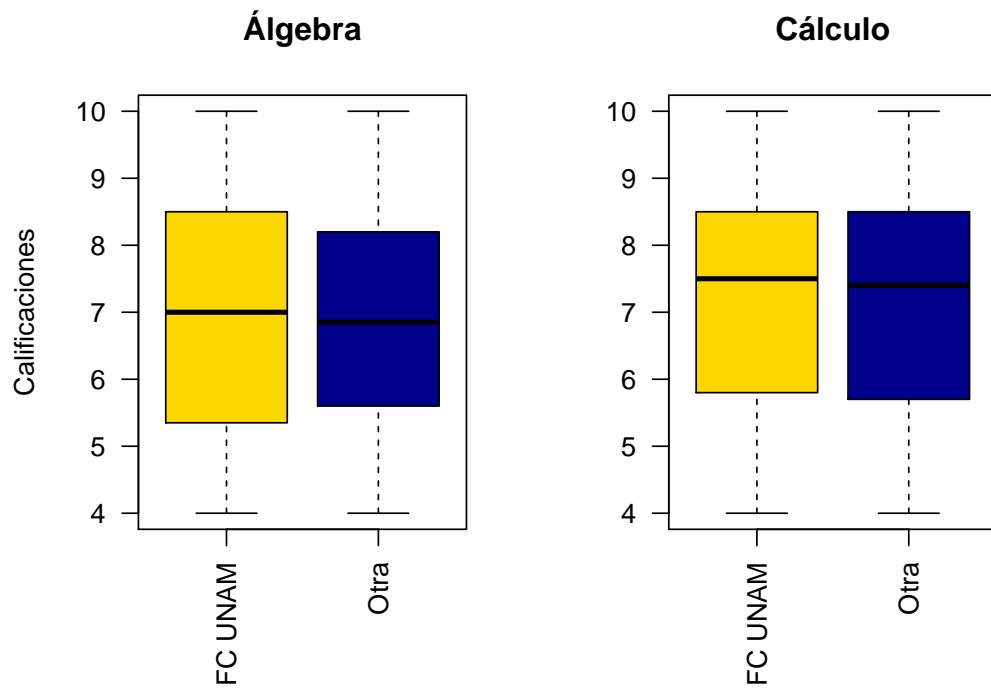


Figura 8: Boxplot de calificaciones de examen por Universidad de los aspirantes.

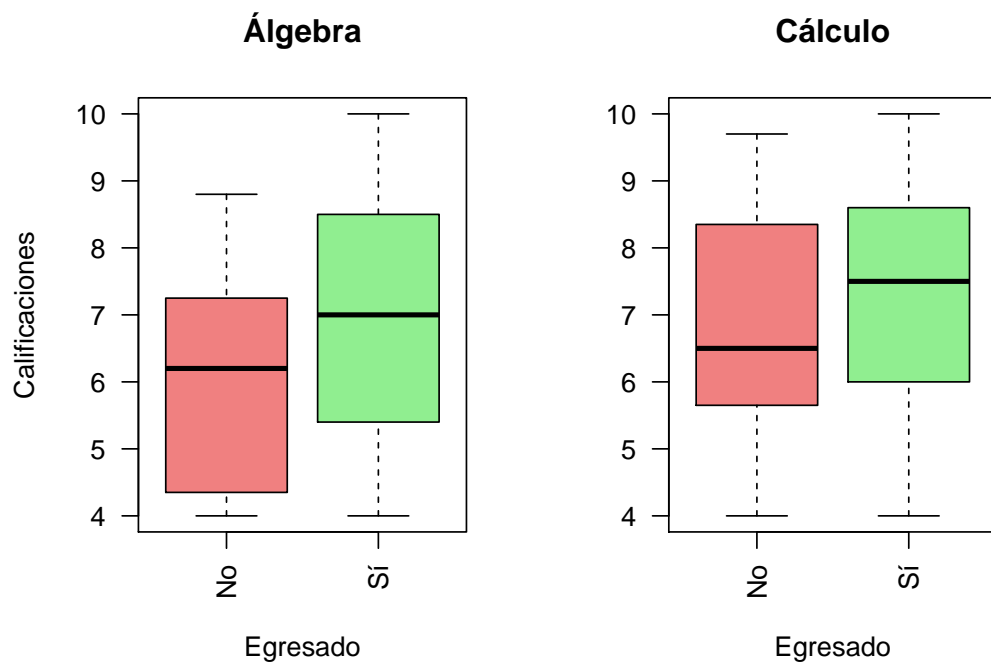


Figura 9: Boxplot de calificaciones de examen por condición de Egreso de los aspirantes.

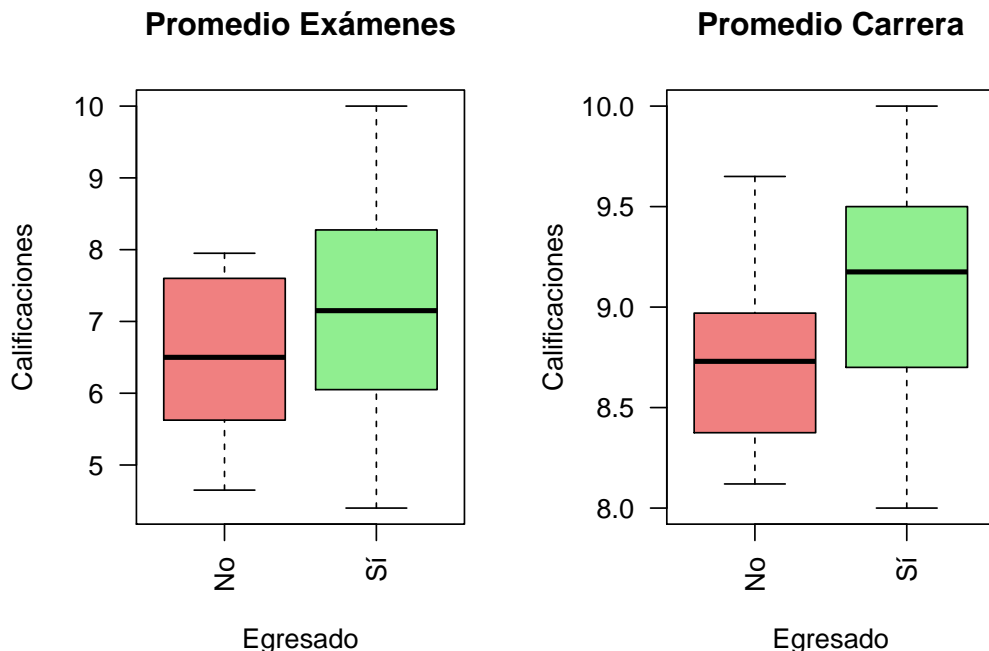


Figura 10: Boxplot de promedio de calificaciones por condición de Egreso de los aspirantes.

cuando un alumno obtiene una calificación muy grave en este examen es bastante probable que no se gradúe de la maestría. Por otro lado, el examen de cálculo a comparación de los egresados, resulta ser una variable aún mejor para distinguir a los individuos aunque teniendo una interpretación similar a la del examen de lineal.

Debido a lo visto y mencionado en la Figura 11, continuamos realizando un promedio entre las calificaciones de los exámenes de ingreso y comparándolo también con el promedio obtenido en la carrera.

Ahora, vamos a ver un gráfico por estatus de egreso de los alumnos en la Figura 12. Aquí, observamos que en efecto la calificación de los alumnos en los exámenes continúa dando una forma de contrastar si un alumno se graduará o no de la maestría. Pero algo aún más importante es que esto se ve reflejado de forma ahora muy similar con el promedio obtenido en la carrera. **Resulta entonces que para graduarse de la maestría sí influye (más que en el egreso) la calificación obtenida en el examen.** Esto al menos de una forma descriptiva

Todo esto nos da una conclusión muy fuerte y es que en efecto, realizar los exámenes para evaluar si un aspirante logrará graduarse es relevante, pero algo igual o quizás más importante es el promedio obtenido en la carrera ya que esto influye tanto en el egreso como en la graduación, herramienta que nos dará buenos indicios de si el aspirante logrará la meta final. En el gráfico que realizaremos en la siguiente sección veremos el comportamiento grupal de las variables numéricas y tomando a consideración el estatus de graduación del alumno.

- **Variables numéricas por estatus de graduación** (Figura 13). Este gráfico nos da la gran oportunidad de estudiar a fondo el comportamiento de las variables de tipo numérico y su relación lineal. En particular podemos notar lo que intuitivamente es también lógico.

Por un lado tenemos la relación lineal más fuerte, que es entre la variable Edad y Tiempo, algo que tiene sentido, de alguna manera estamos viendo que a mayor edad tenga el aspirante, es porque también tardó más tiempo en terminar la carrera. Nótese que el histograma de Edad, nos dice que la mayoría de los aspirantes tienen entre 20 y 25 años; además, casi todos tardan entre 3 a 5 años en terminar la carrera. La siguiente relación lineal fuerte e interesante que se aprecia se da con la variable Promedio. Notamos que, para las variables Edad y Tiempo, el promedio de la licenciatura guarda una relación lineal negativa. Esto se puede traducir a que los estudiantes más jóvenes en entrar a la maestría tienden a tener mayores calificaciones y terminar más rápido la carrera. Algo que también se ve reflejado en los exámenes de admisión, pues con estos se guarda una considerable relación lineal positiva, lo cual favorece en que los estudiantes con

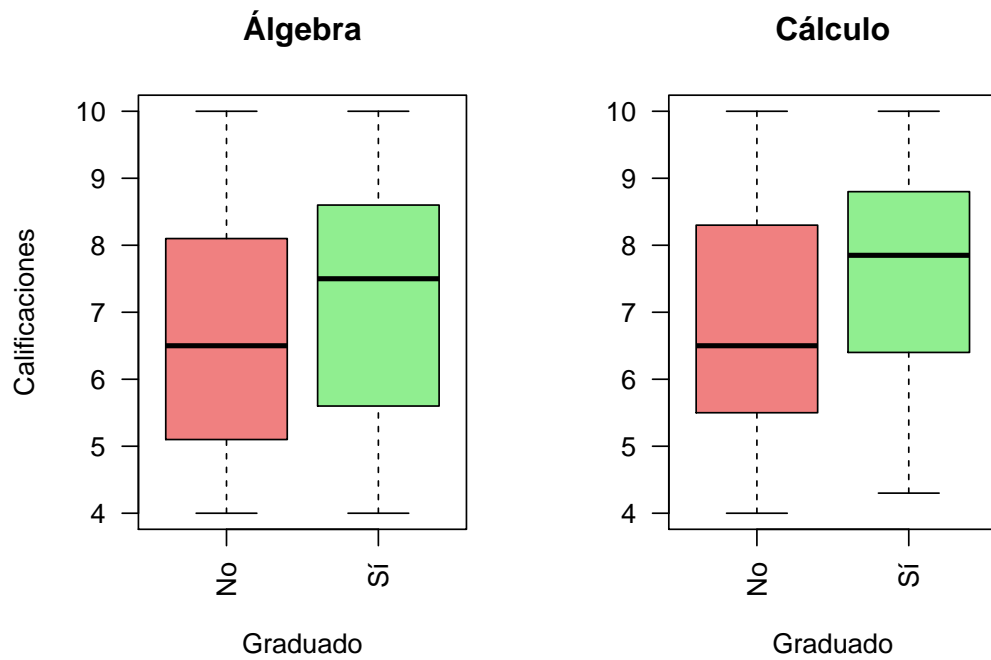


Figura 11: Boxplot de calificaciones de examen por condición de Graduado de los aspirantes.

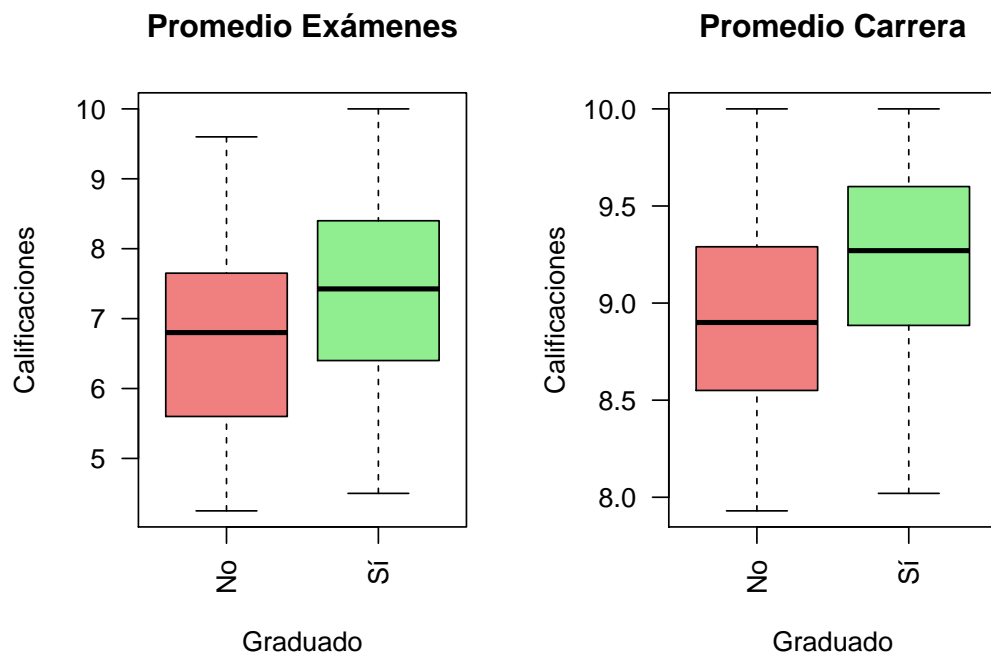


Figura 12: Boxplot de promedio de calificaciones por condición de Graduado de los aspirantes.

buen promedio de la carrera, evidentemente tienden a tener mejores calificaciones en los exámenes de ingreso.

La última relación interesante que mencionaremos se da precisamente entre las calificaciones de los exámenes de ingreso, donde podemos apreciar que tienen una relación lineal positiva, por lo que, los alumnos que usualmente tienen buena calificación en un examen, en el otro también la tendrán. En cuanto a la relación con la variable respuesta, vemos que no es del todo distinguible algún patrón o comportamiento, salvo quizás, en la parte de calificaciones como el promedio y los exámenes de admisión, pues se observa ligeramente que los de valores altos parece que logran graduarse (triángulos amarillos) mientras que los que tienen bajos estos valores parecen no graduarse (triángulos verdes).

Hay que tener cuidado en que, las interpretaciones dadas anteriormente están en términos de una relación lineal que estén guardando las variables, esto nos puede servir también para ver si hay indicios de multicolinealidad entre las variables explicativas del modelo. Vale la pena mencionar, que en el la Figura 13, en la parte diagonal inferior de dicho gráfico, podemos ver a los individuos graficados variable contra variable en cuestión. Al mismo tiempo, podemos ver una pequeña regresión lineal con intervalo de confianza, que nos permite ver más a profundidad la posible relación lineal entre estas variables.

De estos últimos gráficos, podemos concluir que, aunque existe relación entre las variables explicativas, parece ser que tenemos buenas herramientas para poder utilizar un modelo lineal generalizado. Esto lo estaremos realizando en la siguiente sección.

Relación entre las variables continuas

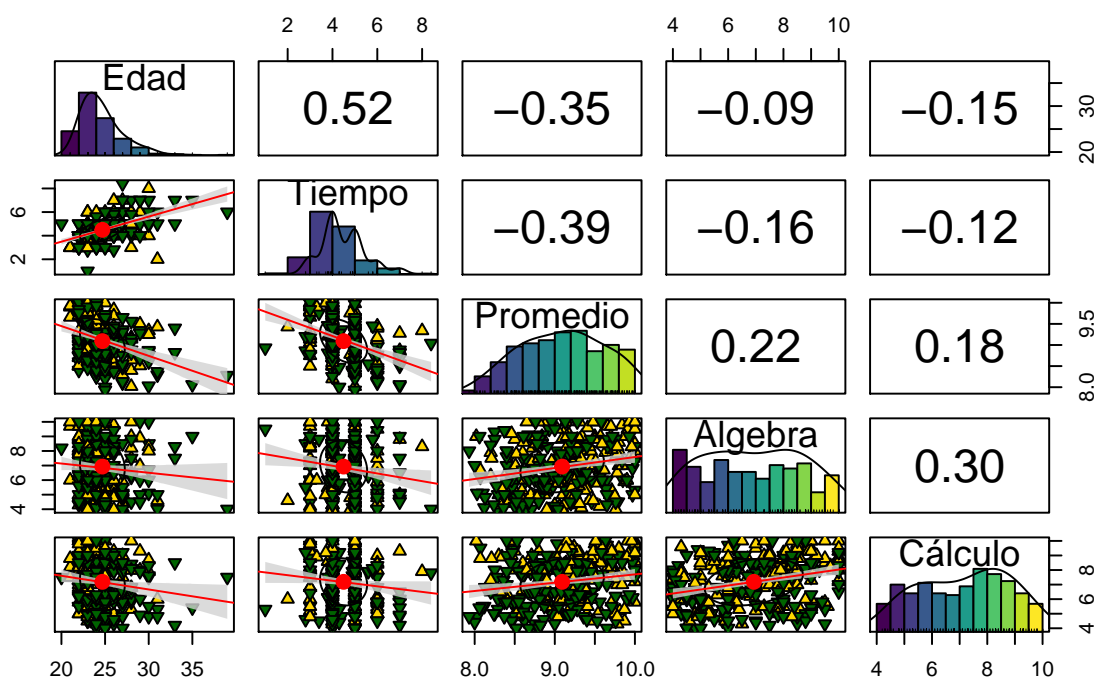


Figura 13: Boxplot de calificaciones de examen por condición de Egreso de los aspirantes.

4. Modelos

Dado que contamos con dos variables respuesta dicotómicas, podemos descartar modelar sus medias con una distribución continua, con la poisson, con una binomial negativa y binomial con parámetro $n \neq 1$. Por lo que, dentro de los modelos lineales generalizados, tenemos dos principales y clásicas propuestas para intentar obtener un modelo. Esto es, usando básicamente la distribución Bernoulli, o bien, $\text{Binomial}(n = 1, p)$ y dos ligas muy populares derivadas de esta propuesta, la liga logística y la probit.

Antes de comenzar, recordemos que en un modelo lineal generalizado, básicamente estamos modelando la media de la respuesta del individuo i , es decir $\mathbb{E}[Y] = \mu_i$ a través de la transformación $g(\cdot)$ (función liga) como la combinación lineal de las covariables del individuo (\underline{x}_i) con los coeficientes (β). Esto es

$$g(\mu_i) = \beta^T \underline{x}_i \Leftrightarrow \mu_i = g^{-1}(\beta^T \underline{x}_i)$$

Y que entonces los coeficientes tendrán interpretación de impacto sobre la media del individuo i con base en dejar el resto de las variables fijas y a través de la inversa de la función liga $g^{-1}(\cdot)$. Por cuestiones entonces de interpretabilidad, nosotros estaremos mostrando los coeficientes transformados en todos nuestros resultados.

Un último detalle a comentar previo a realizar propuestas de modelos es que, recordemos que del análisis exploratorio de los datos, nos enfrentamos a una dificultad interesante y es que los datos no están balanceados. En otras palabras, derivado de que tenemos muchos datos faltantes y más aún, pocos individuos con respuesta de estatus de egreso como “No” entonces tendremos que tener cuidado al momento de sacar o meter variables. Esto tiene un impacto grande ya que, de no tener cuidado, podríamos quedarnos con un modelo cuya variable respuesta consista únicamente en individuos que tengan como respuesta “Sí”, tal es el caso de la variable Fecha_Examen que no cuenta con individuos diferentes de “NA” tales que su estatus de Egresado sea “No”.

```
table(Egresado[!is.na(Fecha_Examen)])
```

```
##
## No  Sí
##  0 156
```

Por consiguiente y derivado de este problema, lo que haremos será proponer 8 modelos diferentes para cada una de las ligas que utilizaremos. Estos modelos nacen del análisis exploratorio, los requisitos del contexto del problema y sentido común. La manera en que vamos a proponer las covariables será la siguiente.

- Modelo 0: Respuesta \sim Sexo. Derivado del análisis descriptivo, este modelo lo proponemos únicamente para destacar que el género de un estudiante **NO** influye en el hecho de si egresará o no de la maestría. Esto es algo que, aunque inicialmente parece influenciar para entrar o no a la maestría, condicionado a que ya se entró, esto no genera influencia significativa.
- Modelo 1: Respuesta \sim Edad+Tiempo. Este modelo lo creamos para ver si aspectos como el tiempo natural de un alumno tiene algún impacto en su respuesta al estatus de Egreso. Es decir, este modelo buscaba explorar la posibilidad de la edad o bien el tiempo que tardaron en realizar su carrera influenciaba de alguna manera o tenía un impacto en la respuesta.
- Modelo 2: Respuesta \sim Promedio+Cálculo+Álgebra. Este modelo fue propuesto para evaluar en aspectos académicos con los que contamos a los individuos en términos de la respuesta. El objetivo es ver si lo que los alumnos presentan académicamente son herramientas suficientes como para lograr el objetivo de egresar de la maestría.
- Modelo 3: Respuesta \sim Universidad+Carrera. Este modelo busca encontrar una influencia sobre los aspectos escolares previos a la maestría. Esto es, si de alguna manera las *alma máter* tienen un impacto sobre si el alumno logrará egresar.
- Modelo 4: Respuesta \sim Veces+Posponer. Aquí estamos buscando la posible relación entre si los alumno que ha tenido una anomalía con el ingreso a la maestría y el hecho de graduarse de la misma. Esto puede ser de interés pues de existir alguna relación, el comité podría considerar otorgar esta clase de oportunidades a los aspirantes.
- Modelo 5: Respuesta \sim Semestre. En particular este modelo se usó para explorar la posible relación de egreso con la fecha en que el alumno entró a la maestría. Esta relación suena un poco vaga en sí misma, pues salvo que por el hecho de que haya habido una generación tal que los profesores dieron particularmente bien o mal las clases o alguna situación de otra naturaleza, entonces se podría explicar la respuesta.
- Modelo 6: Respuesta \sim Promedio+Promedio_Ex. Este es uno de los más interesantes, pues de hecho estamos usando las variables explicativas que se usan en posgrado para dar pie a un aspirante a entrar a la maestría. De tal manera que resulta de interés fundamental, ya que esto es lo que da entrada a un alumno a la maestría y claro, el

objetivo final es que los alumnos que ingresen logren culminar sus estudios, teniendo así un impacto sobre la respuesta.

- Modelo 7: Respuesta \sim Promedio. Este último modelo lo seleccionamos de la parte de estadística descriptiva, ya que vemos que hay una diferencia notoria entre los alumnos que logran egresar con el promedio obtenido en la carrera. Es un modelo que además, guarda parsimonia con respecto al modelo anterior.

Nota: Más modelos fueron utilizados, tratando de hacer combinaciones de estas variables, sin embargo, los resultados obtenidos dada la información con la que contamos no fueron del todo relevantes, así que, se expondrá lo que más logra explicar las variables y hace sentido tomar a consideración.

Lo anterior lo realizaremos primero para ambas variables respuesta

4.1 Para los Egresados

4.1.1 Regresión Logística / Logit

Comenzamos con el modelo más usual y clásico que podemos aplicar a este tipo de comportamientos de variable respuesta. Para esto, recordemos que la inversa de la función logit es

$$g^{-1}(x) = \frac{1}{1 + e^{-x}}$$

Que en código viene dada por

```
inv.liga <- function(x){
  plogis(x)
}
```

- Modelo 0 (Logit) - Tabla 3

```
fit0logit<-glm(Egresado~Sexo,family = binomial(link = "logit"))
```

Tabla 3: Modelo 0 para la liga Logit de los Egresados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0.972	(0.883,0.998)	0	***
SexoHombre	0.285	(0.021,0.675)	0.381	

De este modelo, podemos ver que el Sexo no parece tener una influencia sobre la variable respuesta. Algo que suena lógico pues no debería ser algo que impida o motive a un alumno a terminar sus estudios.

- Modelo 1 (Logit) - Tabla 4

```
fit1logit<-glm(Egresado~Edad+Tiempo,family = binomial(link = "logit"))
```

Tabla 4: Modelo 1 para la liga Logit de los Egresados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	1	(0.93,1)	0.002	**
Edad	0.453	(0.399,0.514)	0.096	.
Tiempo	0.482	(0.337,0.631)	0.811	

Vemos que la edad tiene una influencia pequeña sobre la variable respuesta, al menos tomándola en conjunto con el tiempo en que logró culminar su carrera. Esto quizás hace sentido ya que, como en muchas ocasiones, la edad puede ser un factor que puede motivar a culminar algún proyecto, o en pocas palabras, a terminar lo comenzado.

■ Modelo 2 (Logit) - Tabla 5

```
fit2logit<-glm(Egresado~Promedio+Cálculo+Álgebra,family = binomial(link = "logit"))
```

Tabla 5: Modelo 2 para la liga Logit de los Egresados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0	(0,0.236)	0.033	*
Promedio	0.789	(0.554,0.926)	0.024	*
Cálculo	0.524	(0.436,0.611)	0.595	
Álgebra	0.541	(0.458,0.628)	0.349	

Podemos observar que lo visto en la estadística descriptiva parece cobrar fuerza aquí. Ya que el promedio de la carrera parece ser una variable que en efecto explica bastante bien la respuesta. Lo interesante es que, al menos por separado, la calificación de los exámenes de admisión no parecen tener una influencia significativa.

■ Modelo 3 (Logit) - Tabla 6

```
fit3logit<-glm(Egresado~Universidad+Carrera,family = binomial(link = "logit"))
```

Tabla 6: Modelo 3 para la liga Logit de los Egresados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0.955	(0.815,0.997)	0.003	**
UniversidadOtra	0.49	(0.185,0.87)	0.963	
CarreraOtras	1	(0,NA)	0.991	
CarreraFísica	0.3	(0.015,0.921)	0.564	
CarreraMatemáticas	0.406	(0.035,0.793)	0.722	
CarreraMat. Aplicadas	0.303	(0.018,0.841)	0.521	

Parece que las variables que representan los antecedentes académicos de los estudiantes en términos de la Carrera y lugar de estudios previo, tampoco muestra una influencia sobre la respuesta. Algo interesante, ya que en este sentido, podemos decir al menos que no es suficiente fijarnos en eso para que digamos que un alumno será egresado.

■ Modelo 4 (Logit) - Tabla 7

```
fit4logit<-glm(Egresado~Veces+Posponer,family = binomial(link = "logit"))
```

Tabla 7: Modelo 4 para la liga Logit de los Egresados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0.945	(0.906,0.971)	0	***
Veces	0.525	(0.321,0.827)	0.849	
Posponer	0.404	(0.242,0.628)	0.345	

En este modelo, tampoco vemos que exista influencia al menos con respecto a estas variables. Esto debe ser algo que el comité académico debería pensar como que es una acción neutral al hecho de que los estudiantes logren el objetivo de egresar.

■ Modelo 5 (Logit) - Tabla 8

```
fit5logit<-glm(Egresado~Semestre,family = binomial(link = "logit"))
```

Como ya lo habíamos pensado desde el previo, vemos que la influencia del semestre no parece afectar a la respuesta.

Tabla 8: Modelo 5 para la liga Logit de los Egresados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0.975	(0.895,0.999)	0	***
Semestre2016-2	0.179	(0.01,0.708)	0.226	
Semestre2015-2	0.271	(0.017,0.802)	0.428	
Semestre2018-2	0.25	(0.015,0.785)	0.38	
Semestre2017-2	0.161	(0.008,0.682)	0.191	
Semestre2015-1	0.443	(0.03,0.954)	0.873	
Semestre2019-1	1	(0,NA)	0.993	
Semestre2018-1	0.297	(0.019,0.822)	0.49	
Semestre2017-1	0.17	(0.01,0.63)	0.181	

■ Modelo 6 (Logit) - Tabla 9

```
fit6logit<-glm(Egresado~Promedio+Promedio_Ex,family = binomial(link = "logit"))
```

Tabla 9: Modelo 6 para la liga Logit de los Egresados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0	(0,0.211)	0.031	*
Promedio	0.79	(0.556,0.926)	0.023	*
Promedio_Ex	0.565	(0.464,0.667)	0.22	

Desde el previo habíamos pensado que este modelo debería ser importante a considerar por el comité, sorprendentemente vemos que el promedio de los exámenes generales parece no importar en conjunto con el promedio de la carrera. Esto es importante, ya que pareciera ser que, de los filtros para el ingreso a la maestría, el único que realmente influye en el hecho de que los alumnos logren ser egresados es precisamente el promedio de la carrera.

■ Modelo 7 (Logit) - Tabla 10

```
fit7logit<-glm(Egresado~Promedio,family = binomial(link = "logit"))
```

Tabla 10: Modelo 7 para la liga Logit de los Egresados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0	(0,0.28)	0.037	*
Promedio	0.812	(0.6,0.933)	0.01	**

Llegamos a este último modelo, donde podemos ver que la variable promedio resulta ser muy significativa para evaluar si las personas lograrán egresar de la maestría. Esto lo podemos interpretar como que, cuando un alumno tiene el interés desde la carrera por aprender más, se lo lleva hasta la maestría donde también logrará sus objetivos con una probabilidad alta.

4.1.2 Regresión Probit

Recordemos que la inversa de la función liga es

$$g^{-1}(x) = \Phi(x)$$

Que en código viene dada por

```
inv.liga <- function(x){
  pnorm(x)
}
```

■ Modelo 0 (Probit) - Tabla 11

```
fit0probit<-glm(Egresado~Sexo,family = binomial(link = "probit"))
```

Tabla 11: Modelo 0 para la liga Probit de los Egresados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0.972	(0.883,0.998)	0	***
SexoHombre	0.339	(0.07,0.64)	0.355	

Así como con la liga anterior, de este modelo, podemos ver que el Sexo no parece tener una influencia sobre la variable respuesta. Algo que suena lógico pues no debería ser algo que impida o motive a un alumno en terminar sus estudios.

■ Modelo 1 (Probit) - Tabla 12

```
fit1probit<-glm(Egresado~Edad+Tiempo,family = binomial(link = "probit"))
```

Tabla 12: Modelo 1 para la liga Probit de los Egresados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	1	(0.94,1)	0.002	**
Edad	0.46	(0.415,0.509)	0.091	.
Tiempo	0.491	(0.371,0.611)	0.879	

También aquí vemos que la edad tiene una influencia pequeña sobre la variable respuesta, al menos tomándola en conjunto con el tiempo en que logró culminar su carrera. Esto quizás hace sentido ya que, como en muchas ocasiones, la edad puede ser un factor que puede motivar a culminar algún proyecto, o en pocas palabras, a terminar lo comenzado.

■ Modelo 2 (Probit) - Tabla 13

```
fit2probit<-glm(Egresado~Promedio+Cálculo+Álgebra,family = binomial(link = "probit"))
```

Tabla 13: Modelo 2 para la liga Probit de los Egresados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0	(0,0.348)	0.04	*
Promedio	0.739	(0.543,0.888)	0.021	*
Cálculo	0.515	(0.447,0.582)	0.679	
Álgebra	0.536	(0.472,0.603)	0.281	

Nuevamente, podemos observar que lo visto en la estadística decriptiva parece cobrar fuerza aquí. Ya que el promedio parece ser una variable que en efecto explica bastante bien la respuesta. Lo interesante es que, al menos por separado, la calificación de los exámenes de admisión no parecen tener una influencia significativa.

■ Modelo 3 (Probit) - Tabla 14

```
fit3probit<-glm(Egresado~Universidad+Carrera,family = binomial(link = "probit"))
```

También aquí parece que las variables que representan los antecedentes académicos de los estudiantes en términos de la Carrera y lugar de estudios previo, tampoco muestra una influencia sobre la respuesta. Algo interesante, ya que en este sentido, podemos decir al menos que no es suficiente fijarnos en eso para que digamos que un alumno será egresado.

Tabla 14: Modelo 3 para la liga Probit de los Egresados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0.955	(0.815,0.997)	0	***
UniversidadOtra	0.495	(0.223,0.806)	0.975	
CarreraOtras	1	(0,NA)	0.99	
CarreraFísica	0.341	(0.032,0.855)	0.566	
CarreraMatemáticas	0.429	(0.095,0.752)	0.715	
CarreraMat. Aplicadas	0.343	(0.043,0.786)	0.513	

■ Modelo 4 (Probit) - Tabla 15

```
fit4probit<-glm(Egresado~Veces+Posponer,family = binomial(link = "probit"))
```

Tabla 15: Modelo 4 para la liga Probit de los Egresados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0.945	(0.906,0.971)	0	***
Veces	0.516	(0.348,0.728)	0.87	
Posponer	0.427	(0.284,0.594)	0.376	

Una vez más, en este modelo, tampoco vemos que exista influencia al menos con respecto a estas variables. Esto debe ser algo que el comité académico debería pensar como que es una acción neutral al hecho de que los estudiantes logren el objetivo de egresar.

■ Modelo 5 (Probit) - Tabla 16

```
fit5probit<-glm(Egresado~Semestre,family = binomial(link = "probit"))
```

Tabla 16: Modelo 5 para la liga Probit de los Egresados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0.975	(0.895,0.999)	0	***
Semestre2016-2	0.24	(0.027,0.655)	0.216	
Semestre2015-2	0.329	(0.052,0.729)	0.42	
Semestre2018-2	0.31	(0.046,0.715)	0.371	
Semestre2017-2	0.22	(0.023,0.636)	0.182	
Semestre2015-1	0.461	(0.089,0.878)	0.873	
Semestre2019-1	1	(0,NA)	0.988	
Semestre2018-1	0.352	(0.06,0.746)	0.484	
Semestre2017-1	0.23	(0.029,0.599)	0.163	

Podemos ver nuevamente que como ya lo habíamos pensado desde el previo, vemos que la influencia del semestre no parece afectar a la respuesta.

■ Modelo 6 (Probit) - Tabla 17

```
fit6probit<-glm(Egresado~Promedio+Promedio_Ex,family = binomial(link = "probit"))
```

Así como con la liga anterior, ya desde el previo habíamos pensado que este modelo debería ser importante a considerar por el comité, sorprendentemente vemos que el promedio de los exámenes generales parece no importar en conjunto con el promedio de la carrera. Esto es importante, ya que pareciera ser que, de los filtros para el ingreso a la maestría, el

Tabla 17: Modelo 6 para la liga Probit de los Egresados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0	(0,0.34)	0.038	*
Promedio	0.738	(0.543,0.887)	0.021	*
Promedio_Ex	0.551	(0.471,0.634)	0.21	

único que realmente influye en el hecho de que los alumnos logren ser egresados es precisamente el promedio de la carrera.

■ Modelo 7 (Probit) - Tabla 18

```
fit7probit<-glm(Egresado~Promedio,family = binomial(link = "probit"))
```

Tabla 18: Modelo 7 para la liga Probit de los Egresados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0	(0,0.395)	0.043	*
Promedio	0.761	(0.579,0.897)	0.008	**

Así, vemos de nuevo que llegamos a este último modelo, donde podemos ver que la variable promedio resulta ser muy significativa para evaluar si las personas lograrán egresar de la maestría. Esto lo podemos interpretar como que, cuando un alumno tiene el interés desde la carrera por aprender más, se lo lleva hasta la maestría donde también logrará sus objetivos con una probabilidad alta.

4.2 Para los Graduados

Para esta sección lo que haremos será quitar de los datos a aquellos alumnos que ingresaron en el semestre más actual, es decir, los que están en el semestre 2019-2 de tal manera que esos de hecho vamos a predecirlos una vez seleccionando el mejor modelo. La razón por la que hacemos esto es porque no sería justo incluir un modelo que los contemple pues **a la fecha de este trabajo** resulta complicado que ya se hayan graduado, pero siendo exigentes, sí pudieron ya haber culminado sus créditos, es decir, que egresaron.

```
test <- datos[datos$Semestre=="2019-2",]
datos <- datos[!datos$Semestre=="2019-2",]
```

4.2.1 Regresión Logística / Logit

Continuamos regresando con el modelo más usual y clásico que podemos aplicar a este tipo de comportamientos de variable respuesta.

■ Modelo 0 (Logit) - Tabla 19

```
fit0logit<-glm(Graduado~Sexo,family = binomial(link = "logit"))
```

Tabla 19: Modelo 0 para la liga Logit de los Graduados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0.619	(0.468,0.756)	0.127	
SexoHombre	0.452	(0.292,0.616)	0.576	

De este modelo, podemos ver que el Sexo no parece tener una influencia sobre la variable respuesta. Algo que suena lógico pues no debería ser algo que impida o motive a un alumno en terminar sus estudios. **Notemos que a diferencia de la respuesta anterior, vemos aquí que ni si quiera el intercepto resulta significativo.**

■ Modelo 1 (Logit) - Tabla 20

```
fit1logit<-glm(Graduado~Edad+Tiempo,family = binomial(link = "logit"))
```

Tabla 20: Modelo 1 para la liga Logit de los Graduados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0.989	(0.858,0.999)	0.002	**
Edad	0.459	(0.425,0.49)	0.012	*
Tiempo	0.477	(0.404,0.55)	0.537	

Vemos que la edad tiene una influencia un poco más grande sobre la variable respuesta, al menos tomándola en conjunto con el tiempo en que logró culminar su carrera. Esto quizás hace sentido ya que, como en muchas ocasiones, la edad puede ser un factor que puede motivar a culminar algún proyecto, o en pocas palabras, a terminar lo comenzado.

■ Modelo 2 (Logit) - Tabla 21

```
fit2logit<-glm(Graduado~Promedio+Cálculo+Álgebra,family = binomial(link = "logit"))
```

Tabla 21: Modelo 2 para la liga Logit de los Graduados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0	(0,0.002)	0	***
Promedio	0.722	(0.615,0.811)	0	***
Cálculo	0.563	(0.524,0.602)	0.002	**
Álgebra	0.509	(0.473,0.545)	0.624	

Podemos observar que lo visto en la estadística descriptiva parece cobrar fuerza aquí. Ya que el promedio parece ser una variable que en efecto explica bastante bien la respuesta. Lo interesante es que, **ahora el examen de cálculo resulta significativo** algo MUY interesante, ya que para los egresados parece que el examen de Álgebra (lineal) era importante, pero ahora resulta que para los graduados el más significativo el de cálculo. **Esto le da importancia y sentido a realizar ambos exámenes.**

■ Modelo 3 (Logit) - Tabla 22

```
fit3logit<-glm(Graduado~Universidad+Carrera,family = binomial(link = "logit"))
```

Tabla 22: Modelo 3 para la liga Logit de los Graduados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0.346	(0.184,0.537)	0.123	
UniversidadOtra	0.493	(0.317,0.673)	0.939	
CarreraOtras	0.39	(0.102,0.749)	0.584	
CarreraFísica	0.694	(0.351,0.909)	0.264	
CarreraMatemáticas	0.719	(0.526,0.863)	0.032	*
CarreraMat. Aplicadas	0.538	(0.275,0.783)	0.792	

Parece que las variables que representan los antecedentes académicos de los estudiantes en términos de la Carrera y lugar de estudios previo, tampoco muestra una influencia sobre la respuesta **salvo en la carrera de matemáticas** la cuál es significativa, algo sumamente interesante, parece que de hecho hay un aumento a favor de graduarse si se viene de esta carrera.

■ Modelo 4 (Logit) - Tabla 23

```
fit4logit<-glm(Graduado~Veces+Posponer,family = binomial(link = "logit"))
```

Tabla 23: Modelo 4 para la liga Logit de los Graduados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0.575	(0.509,0.639)	0.027	*
Veces	0.426	(0.329,0.525)	0.148	
Posponer	0.342	(0.246,0.447)	0.005	**

Este modelo vaya que es interesante, pues estamos viendo que el hecho de posponer la entrada a la maestría es un factor que influye a que los alumnos **NO** se titulen de manera significativa, algo muy importante a considerar por parte del comité.

■ Modelo 5 (Logit) - Tabla 24

```
fit5logit<-glm(Graduado~Semestre,family = binomial(link = "logit"))
```

Tabla 24: Modelo 5 para la liga Logit de los Graduados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0	(0,1)	0.98	
Semestre2016-1	1	(1,1)	0.979	
Semestre2016-2	1	(1,1)	0.979	
Semestre2015-2	1	(0.011,1)	0.979	
Semestre2018-2	1	(0.003,1)	0.981	
Semestre2017-2	1	(0.011,1)	0.979	
Semestre2015-1	1	(1,1)	0.978	
Semestre2019-1	1	(0.001,1)	0.981	
Semestre2018-1	1	(0.011,1)	0.98	
Semestre2017-1	1	(0.997,1)	0.979	

Como ya lo habíamos pensado desde el previo, vemos que la influencia del semestre no parece afectar a la respuesta.

■ Modelo 6 (Logit) - Tabla 25

```
fit6logit<-glm(Graduado~Promedio+Promedio_Ex,family = binomial(link = "logit"))
```

Tabla 25: Modelo 6 para la liga Logit de los Graduados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0	(0,0.003)	0	***
Promedio	0.716	(0.609,0.806)	0	***
Promedio_Ex	0.568	(0.523,0.614)	0.003	**

Desde el previo habíamos pensado que este modelo debería ser importante a considerar por el comité, sorprendentemente vemos que el promedio de los exámenes generales parece ser significativo **para la graduación del alumno** en conjunto con el promedio de la carrera. Esto es importante, ya que pareciera ser que, de los filtros para el ingreso a la maestría, el único que realmente influye en el hecho de que los alumnos logren ser egresados es precisamente el promedio de la carrera. **Pero para que los alumnos alcancen la meta final de titulación parece importar significativamente el promedio de los exámenes de admisión.**

■ Modelo 7 (Logit) - Tabla 26

```
fit7logit<-glm(Graduado~Promedio,family = binomial(link = "logit"))
```

Tabla 26: Modelo 7 para la liga Logit de los Graduados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0	(0,0.004)	0	***
Promedio	0.745	(0.648,0.826)	0	***

Llegamos a este último modelo, donde podemos ver que la variable promedio resulta ser muy significativa para evaluar si las personas lograrán graduarse de la maestría. Esto lo podemos interpretar como que, cuando un alumno tiene el interés desde la carrera por aprender más, se lo lleva hasta la maestría donde también logrará sus objetivos con una probabilidad alta.

4.1.2 Regresión Probit

- Modelo 0 (Probit) - Tabla 27

```
fit0probit<-glm(Graduado~Sexo,family = binomial(link = "probit"))
```

Tabla 27: Modelo 0 para la liga Probit de los Graduados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0.619	(0.468,0.756)	0.123	
SexoHombre	0.452	(0.294,0.617)	0.575	

Así como con la liga anterior, de este modelo, podemos ver que el Sexo no parece tener una influencia sobre la variable respuesta. Algo que suena lógico pues no debería ser algo que impida o motive a un alumno en terminar sus estudios. **Notemos que a diferencia de la respuesta anterior, vemos aquí que ni si quiera el intercepto resulta significativo**

- Modelo 1 (Probit) - Tabla 28

```
fit1probit<-glm(Graduado~Edad+Tiempo,family = binomial(link = "probit"))
```

Tabla 28: Modelo 1 para la liga Probit de los Graduados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0.997	(0.873,1)	0.001	**
Edad	0.459	(0.427,0.489)	0.01	**
Tiempo	0.477	(0.407,0.547)	0.53	

También aquí vemos que la edad tiene una influencia ya muy buena influencia sobre la variable respuesta, al menos tomándola en conjunto con el tiempo en que logró culminar su carrera. Esto quizás hace sentido ya que, como en muchas ocasiones, la edad puede ser un factor que puede motivar a culminar algún proyecto, o en pocas palabras, a terminar lo comenzado.

- Modelo 2 (Probit) - Tabla 29

```
fit2probit<-glm(Graduado~Promedio+Cálculo+Álgebra,family = binomial(link = "probit"))
```

Nuevamente, podemos observar que lo visto en la estadística descriptiva parece cobrar fuerza aquí. Ya que el promedio de la carrera parece ser una variable que en efecto explica bastante bien la respuesta. Lo interesante es que, **ahora el examen de cálculo resulta significativo** algo MUY interesante, ya que para los egresados parece que el examen de Álgebra (lineal) era importante, pero ahora resulta que para los graduados el más significativo el de cálculo. **Esto le da importancia y sentido a realizar ambos exámenes.**

Tabla 29: Modelo 2 para la liga Probit de los Graduados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0	(0,0)	0	***
Promedio	0.721	(0.614,0.812)	0	***
Cálculo	0.562	(0.524,0.6)	0.002	**
Álgebra	0.509	(0.474,0.545)	0.602	

■ Modelo 3 (Probit) - Tabla 30

```
fit3probit<-glm(Graduado~Universidad+Carrera,family = binomial(link = "probit"))
```

Tabla 30: Modelo 3 para la liga Probit de los Graduados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0.346	(0.184,0.537)	0.118	
UniversidadOtra	0.494	(0.318,0.671)	0.947	
CarreraOtras	0.394	(0.106,0.748)	0.58	
CarreraFísica	0.695	(0.352,0.921)	0.263	
CarreraMatemáticas	0.721	(0.526,0.869)	0.029	*
CarreraMat. Aplicadas	0.537	(0.276,0.783)	0.793	

También aquí parece que las variables que representan los antecedentes académicos de los estudiantes en términos de la Carrera y lugar de estudios previo, tampoco muestra una influencia sobre la respuesta **salvo en la carrera de matemáticas** la cuál es significativa, algo sumamente interesante, parece que de hecho hay un aumento a favor de graduarse si se viene de esta carrera.

■ Modelo 4 (Probit) - Tabla 31

```
fit4probit<-glm(Graduado~Veces+Posponer,family = binomial(link = "probit"))
```

Tabla 31: Modelo 4 para la liga Probit de los Graduados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0.575	(0.509,0.639)	0.027	*
Veces	0.426	(0.33,0.524)	0.142	
Posponer	0.343	(0.247,0.447)	0.004	**

Este modelo vaya que es interesante, pues estamos viendo que el hecho de posponer la entrada a la maestría es un factor que influye a que los alumnos **NO** se titulen de manera significativa, algo muy importante a considerar por parte del comité.

■ Modelo 5 (Probit) - Tabla 32

```
fit5probit<-glm(Graduado~Semestre,family = binomial(link = "probit"))
```

Podemos ver nuevamente que como ya lo habíamos pensado desde el previo, vemos que la influencia del semestre no parece afectar a la respuesta.

■ Modelo 6 (Probit) - Tabla 33

```
fit6probit<-glm(Graduado~Promedio+Promedio_Ex,family = binomial(link = "probit"))
```

Nuevamente, comentamos que desde el previo habíamos pensado que este modelo debería ser importante a considerar por el comité, sorprendentemente vemos que el promedio de los exámenes generales parece ser significativo **para la**

Tabla 32: Modelo 5 para la liga Probit de los Graduados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0	(0,0.997)	0.974	
Semestre2016-1	1	(0.008,1)	0.973	
Semestre2016-2	1	(0.467,1)	0.973	
Semestre2015-2	1	(0.019,1)	0.971	
Semestre2018-2	1	(0.293,1)	0.976	
Semestre2017-2	1	(1,1)	0.972	
Semestre2015-1	1	(0.058,1)	0.969	
Semestre2019-1	1	(0.225,1)	0.977	
Semestre2018-1	1	(0.005,1)	0.973	
Semestre2017-1	1	(0.008,1)	0.973	

Tabla 33: Modelo 6 para la liga Probit de los Graduados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0	(0,0)	0	***
Promedio	0.716	(0.609,0.808)	0	***
Promedio_Ex	0.568	(0.524,0.612)	0.003	**

graduación del alumno en conjunto con el promedio de la carrera. Esto es importante, ya que pareciera ser que, de los filtros para el ingreso a la maestría, el único que realmente influye en el hecho de que los alumnos logren ser egresados es precisamente el promedio de la carrera. **Pero para que los alumnos alcancen la meta final de titulación parece importar significativamente el promedio de los exámenes de admisión.**

- Modelo 7 (Probit) - Tabla 34

```
fit7probit<-glm(Graduado~Promedio,family = binomial(link = "probit"))
```

Tabla 34: Modelo 7 para la liga Probit de los Graduados

	inv.liga(coef)	CI(95 %)	p-value	Signif.
(Intercept)	0	(0,0)	0	***
Promedio	0.746	(0.647,0.83)	0	***

Así, vemos de nuevo que llegamos a este último modelo, donde podemos ver que la variable promedio resulta ser muy significativa para evaluar si las personas lograrán graduarse de la maestría. Esto lo podemos interpretar como que, cuando un alumno tiene el interés desde la carrera por aprender más, se lo lleva hasta la maestría donde también logrará sus objetivos con una probabilidad alta.

5. Selección de Modelo

5.1 Comparativas de Modelo

5.1.1 Para los egresos

Para poder seleccionar un modelo final, consideraremos las estadísticas AIC y BIC para lograr comparar los modelos, esto lo hacemos en la Tabla 35.

Donde podemos seleccionar dos candidatos a modelo final, ambos pertenecientes a la liga Probit, los modelos 1 y 7. Los cuales son seleccionados, uno por menor nivel en las estadísticas en cuestión y otro por mayor representabilidad,

Tabla 35: Comparación de modelos con AICs y BICs.

	AICs	BICs
fit0logit	115.87128	122.87380
fit1logit	80.57410	90.28343
fit2logit	110.94876	124.85461
fit3logit	120.87576	141.50424
fit4logit	118.02260	128.52637
fit5logit	124.76495	156.27627
fit6logit	109.00809	119.43748
fit7logit	108.57023	115.52316
fit0probit	115.87128	122.87380
fit1probit	80.34694	90.05627
fit2probit	110.76285	124.66870
fit3probit	120.87695	141.50542
fit4probit	118.05675	128.56053
fit5probit	124.76495	156.27627
fit6probit	108.92928	119.35867
fit7probit	108.46198	115.41490

sentido lógico y significancia respectivamente.

5.1.2 Para los graduados

Para poder seleccionar un modelo final, consideraremos nuevamente las estadísticas AIC y BIC para lograr comparar los modelos, esto lo hacemos en la Tabla 36.

Tabla 36: Comparación de modelos con AICs y BICs.

	AICs	BICs
fit0logit	369.6948	376.8842
fit1logit	330.6102	341.0894
fit2logit	382.1711	396.9054
fit3logit	394.9891	416.9040
fit4logit	412.2496	423.3709
fit5logit	341.5964	378.6675
fit6logit	383.2901	394.3408
fit7logit	390.1209	397.4880
fit0probit	369.6948	376.8842
fit1probit	330.5173	340.9965
fit2probit	382.0823	396.8166
fit3probit	394.9906	416.9055
fit4probit	412.2336	423.3550
fit5probit	341.5964	378.6675
fit6probit	383.1447	394.1954
fit7probit	390.1745	397.5417

Aquí hay que tener mucho cuidado para proponer el modelo, ya que vamos a seleccionar dos candidatos a modelo final, ambos pertenecientes a la liga Probit, los modelos 1 y 6. Los cuales son seleccionados, uno por menor nivel en las estadísticas en cuestión y otro por mayor representabilidad, sentido lógico y significancia respectivamente. Aunque

podríamos ver que hay otras propuestas de modelos con base únicamente en el AIC y BIC, pero si nos remitimos a lo que estos representarían podría carecer de sentido.

5.2 Interpretación del modelo final

5.2.1 Para los egresados

- Modelo 1 (Probit) - Tabla 12: En términos de los coeficientes, parece que la edad y el tiempo son efectos que a medida que aumentan perjudican la posibilidad de que el alumno no logre egresar de la maestría. Esto se concluye ya que su estimación bajo la inversa de la liga probit es menor a 0.5. Este puede que sea un ejemplo donde considerar el AIC y BIC únicamente no sea siempre del todo apropiado, hay que también buscar un contexto al problema y parece que este es el caso para este modelo, por lo tanto, aunque es un modelo seleccionado, posiblemente el mejor sea el que comentaremos a continuación.
- Modelo 7 (Probit) - Tabla 18: Este es el modelo que más sentido puede tener, y es que parece que el promedio es algo que influye altamente en que si los alumnos lograrán el objetivo de egresar, a medida que el promedio crece vemos que las probabilidades de que el alumno egrese aumentan, de tal manera que podemos decir que este es el efecto más importante que debemos considerar para poder predecir si el alumno logrará el objetivo.

5.2.2 Para los graduados

- Modelo 1 (Probit) - Tabla 28: En términos de los coeficientes, parece que la edad (principalmente) y el tiempo son efectos que a medida que aumentan perjudican la posibilidad de que el alumno no logre graduarse de la maestría. Esto se concluye ya que su estimación bajo la inversa de la liga probit es menor a 0.5. Este puede que sea otro ejemplo donde considerar el AIC y BIC únicamente no sea siempre del todo apropiado, hay que también buscar un contexto al problema y parece que este es el caso para este modelo, por lo tanto, aunque es un modelo seleccionado, posiblemente el mejor sea el que comentaremos a continuación.
- Modelo 6 (Probit) - Tabla 34: Este es el modelo que más sentido puede tener, y es que parece que el promedio **en conjunto con el promedio de los exámenes de admisión** es algo que influye altamente en que si los alumnos lograrán el objetivo de egresar, a medida que los promedios crecen vemos que las probabilidades de que el alumno se gradúe aumentan, de tal manera que podemos decir que estos son los efectos más importantes que debemos considerar para poder predecir si el alumno logrará el objetivo final.

6. Predicciones

En particular las predicciones como ya lo habíamos comentado las haremos sobre los graduados. Estas predicciones las veremos en la tabla de a continuación y se harán con el Modelo 6 del tipo Probit como se seleccionó anteriormente.

```
test$Graduado <- predict(object = fits$fit6probit,
                        newdata = test[,c("Promedio_Ex", "Promedio")]) %>%
  inv.liga() %>%
  round(3)
```

Donde mostramos en la Tabla 37 las probabilidades de graduación estimadas, junto con otras medidas interesantes de los individuos

7. Conclusiones

De todos los modelos estudiados, podemos ver desde la parte descriptiva, que el efecto más poderoso que tenemos para ver si un alumno/aspirante logrará egresar de la maestría es el promedio obtenido en la carrera. **Mientras que para que el alumno logre graduarse, hay que enfocarnos tanto en los exámenes de admisión como en el promedio de la carrera.** Esto tiene mucho sentido y más en la dirección que apunta el modelo final propuesto, a medida que el/los promedio(s) crece podemos pensar que este alumno tiene aspiraciones de complementar su formación académica y por lo tanto logrará los objetivos más importantes de entrar a la maestría con una probabilidad mayor a los que no.

Tabla 37: Predicción de los alumnos graduados para el semestre 2019-2 utilizando el Modelo 6 - Probit.

	Posponer	Edad	Carrera	Universidad	Tiempo	Veces	Promedio_Ex	Promedio	Graduado
270	0	23	Mat. Aplicadas	FC UNAM	4	0	6.85	9.20	0.526
271	0	23	Matemáticas	FC UNAM	5	0	7.05	9.55	0.618
272	0	30	Matemáticas	FC UNAM	6	0	9.15	8.76	0.583
273	0	29	Matemáticas	FC UNAM	5	0	8.95	9.28	0.681
274	1	24	Matemáticas	FC UNAM	3	0	8.25	8.78	0.526
275	0	27	Matemáticas	FC UNAM	7	2	5.40	8.55	0.290
276	0	35	Matemáticas	FC UNAM	7	0	7.20	8.06	0.299
277	1	23	Matemáticas	FC UNAM	4	0	6.00	9.90	0.626
278	1	31	Actuaría	FC UNAM	5	0	4.90	8.37	0.229
279	0	24	Matemáticas	FC UNAM	4	2	7.05	8.61	0.406
280	1	22	Matemáticas	FC UNAM	3	0	7.90	9.84	0.730
281	1	22	Matemáticas	FC UNAM	4	0	5.35	9.76	0.551
282	0	26	Matemáticas	FC UNAM	4	0	7.55	9.28	0.592
283	0	25	Matemáticas	FC UNAM	3	0	6.70	9.18	0.511
284	0	25	Física	FC UNAM	5	0	6.40	7.93	0.230
285	1	30	Matemáticas	FC UNAM	5	0	5.75	8.87	0.378
286	0	29	Matemáticas	FC UNAM	7	0	4.70	8.68	0.274
287	0	24	Mat. Aplicadas	FC UNAM	4	2	7.00	9.21	0.539
288	0	25	Otras	FC UNAM	5	0	9.60	9.58	0.775
289	0	24	Actuaría	FC UNAM	4	2	5.50	9.27	0.450
290	0	NA	NA	NA	NA	0	NA	NA	NA
291	0	24	Matemáticas	FC UNAM	4	0	7.65	9.93	0.733
292	0	25	Matemáticas	Otra	4	0	8.75	9.72	0.755
293	0	25	Mat. Aplicadas	FC UNAM	6	0	6.90	8.40	0.351
294	1	25	Matemáticas	FC UNAM	5	0	4.25	8.52	0.221
295	1	24	Matemáticas	FC UNAM	4	0	9.25	8.78	0.594
296	0	24	Matemáticas	FC UNAM	4	0	6.90	9.00	0.484
297	0	26	Matemáticas	FC UNAM	4	2	6.65	9.31	0.538
298	0	26	Mat. Aplicadas	FC UNAM	4	0	5.55	9.40	0.483
299	1	27	Matemáticas	FC UNAM	4	0	6.65	8.75	0.411
300	1	25	Matemáticas	FC UNAM	4	0	5.65	9.65	0.547
301	0	33	Otras	FC UNAM	7	0	5.85	8.82	0.373

Esto en general nos dice que los métodos propuestos por el comité son esencialmente correctos y buenos filtros para aquellos alumnos que sí están interesados en extender sus conocimientos y grado académico. Quizás una buena sugerencia, sería considerar más a fondo, con base en el modelo y la estadística descriptiva el promedio de los aspirantes para así tener aún más egresados.