

Laboratorio 5

Instrucciones

Este laboratorio hace referencia a temas abordados en clase sobre regresión logística. Tienes la opción de contestarlo en equipos de hasta 3 personas (máx.) o de manera individual. El archivo con las respuestas lo debes enviar por email a mi correo electrónico (sergio.bejar@cide.edu) a más tardar el sábado 25 de mayo, 2023 a las 12:00pm.

La base de datos que vamos a utilizar se encuentra en el folder “Files” del repositorio Github de la clase y se llama (“admission.csv”).

Ejercicio: Determinantes de Admisión a Programa de Posgrado

En este ejercicio analizaremos que variables influyen en la decisión de programas de posgrado en EEUU de admitir a un estudiante o no. En particular, vamos a asumir que la probabilidad de ser admitido a un programa de posgrado (**admit**) está dada en función de la calificación obtenida por el solicitante en el GRE (**gre**), el promedio de calificaciones en la licenciatura (**gpa**) y el ránking de la universidad en cuestión (**rank**).

1. Estima un modelo logit de la especificación descrita.

```
library(foreign)
admission <- read.csv("/Users/sergiobejar/Downloads/admission.csv") # llama database
```

Una vez cargada la base estimamos el modelo de la siguiente forma:

```
m_lab5 <- glm(admit ~ gre + gpa + rank,
              family = binomial(link = "logit"),
              data = admission) # estima modelo logit

library(stargazer)
stargazer(m_lab5, type = "text", covariate.labels = c("Score GRE",
"Promedio licenciatura",
```

```
"Ranking universidad")) #imprime una tabla con los resultados
```

```
=====
                        Dependent variable:
                        -----
                                admit
                        -----
Score GRE                0.002**
                        (0.001)

Promedio licenciatura    0.777**
                        (0.327)

Ranking universidad      -0.560***
                        (0.127)

Constant                -3.450***
                        (1.133)

-----
Observations             400
Log Likelihood            -229.721
Akaike Inf. Crit.         467.442
=====
Note:                    *p<0.1; **p<0.05; ***p<0.01
```

2. Interpreta los coeficientes obtenidos al estimar el modelo en (1).

RESPUESTA: Los resultados indican que hay un efecto positivo y estadísticamente significativo (al 95% de confianza) de las variables **gre** y **gpa** en la probabilidad de ser admitido a un posgrado en EEUU. En contraste, la variable **rank** tiene un efecto negativo sobre nuestra variable dependiente y dicho efecto es también estadísticamente significativo al 95% de confianza.

Es importante recordar que esta es solo una interpretación general de los resultados. Para hacer una inferencia más substantiva tenemos que estimar Odds Ratio y/o probabilidades. Eso lo vamos a hacer a continuación.

3. Calcula e interpreta la razón de momios (Odds Ratio) de los coeficientes de tu modelo.

```
exp(coef(m_lab5)) # estima Odds Ratios para todos los coeficientes
```

(Intercept)	gre	gpa	rank
0.03175998	1.00229659	2.17496718	0.57119114

Una vez que estimamos la razón de momios (Odds Ratio) podemos decir lo siguiente:

- Por una unidad de incremento en el **gre** los momios de ser aceptado en un programa de posgrado (versus no ser aceptado) se incrementan en un factor de 1.002.
- Por una unidad de incremento en el promedio de la licenciatura (**gpa**) los momios de ser admitido en posgrado se incrementan en un factor de 2.174.
- Por cada incremento de una unidad en el ranking de la universidad (**rank**), los momios de ser admitido al posgrado en esa universidad se reducen en un factor de 0.571.

4. Calcula e interpreta las probabilidades de los coeficientes de tu modelo.

```
plogis(coef(m_lab5))
```

(Intercept)	gre	gpa	rank
0.03078233	0.50057349	0.68503611	0.36354020

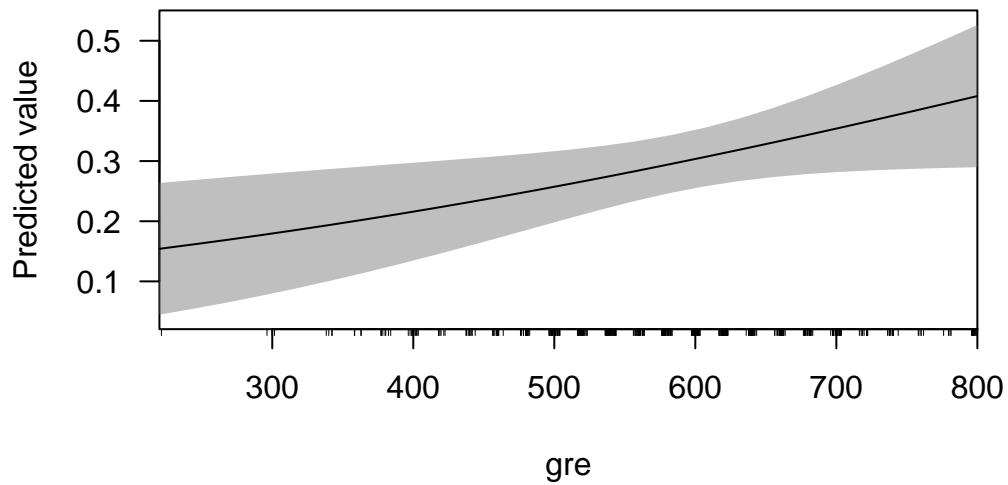
En términos de probabilidad la interpretación es la siguiente:

- Un incremento de una unidad en GRE incrementa la probabilidad de ser aceptado al posgrado en un 50%.
- Un incremento de una unidad en GPA incrementa la probabilidad de ser aceptado al posgrado en un 68%.
- Un incremento de una unidad en RANKING la probabilidad de ser admitido se incrementa un 36.3%.

5. Grafica el efecto de la variable **gre** en la variable dependiente.

Para graficar el efecto de alguna variable predictiva en nuestra variable dependiente tenemos que usar la función **predict** que nos da las probabilidades de que $y=1$ para diferentes valores (i.e. rango) de nuestra variable predictiva. En el documento RPubS con el ejemplo de modelos logit hay dos ejemplos con código para poder hacer este gráfico. En clase mencioné que hay paquetes que nos ayudan a hacer estos gráficos de manera más sencilla y acá va un ejemplo con el paquete **margins**.

```
library(margins) # cargamos paquete margins
cplot(m_lab5, x = "gre", se.type = "shade") # grafica probabilidades predichas
```



6. Evalúa tu modelo usando una prueba Likelihood Ratio.

Esto se puede hacer con una prueba de diferencia de residuos o más directamente con ANOVA

```
anova(m_lab5, test = "Chisq")
```

Analysis of Deviance Table

Model: binomial, link: logit

Response: admit

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)	
NULL			399	499.98		
gre	1	13.9204	398	486.06	0.0001907	***
gpa	1	5.7122	397	480.34	0.0168478	*
rank	1	20.9022	396	459.44	4.833e-06	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Los resultados indican que todos los predictores contribuyen significativamente a explicar la probabilidad admisión a un programa de posgrado (p-values menores a .05 para coeficientes GRE, GPA y RANK).

7. ¿Cuál es la precisión del modelo estimado en (1)?

```
library(vcd)
```

Loading required package: grid

```
predicciones <- ifelse(test = m_lab5$fitted.values > 0.5, yes = 1, no = 0)
matriz_confusion <- table(m_lab5$model$admit, predicciones,
                          dnn = c("observaciones", "predicciones"))
matriz_confusion
```

	predicciones	
observaciones	0	1
0	253	20
1	98	29

El porcentaje de predicciones correctas de nuestro modelo es: $(253+29)/(253+29+20+98) = 0.705 = 70.5\%$.