

## Sesión 3

### Contents

MODELO 1 : Retorno y personas de la tercera edad	3
MODELO 2	4
MODELO 3:	8



### FACULTAD DE CIENCIAS SOCIALES - PUCP

Curso: POL 304 - Estadística para el análisis político 2 | Semestre 2023 - 2

Jefas de Práctica: Karina Alcántara y Lizette Crispín

### SESIÓN 3 - Regresión Logística Binaria

La base que usaremos hoy es la Encuesta Nacional a Docentes de Instituciones Educativas Públicas de Educación Básica Regular



Esta base de datos es del 2020, es decir, que hay que tomar en cuenta que se realizó en contexto de la pandemia. Entonces, hay diversas variables. Con respecto al cuidado de parientes, qué enfermedades ha tenido, satisfacción sobre temas personales o de la misma institución educativa.

```
library(rio)
endo=import("END01.sav")
```

```
library(dplyr)
```

Estas son las variables que usaremos:

**Variable dependiente:** (*P2\_2*) Retorno a clases

**Variables independientes:**

- **P1\_24\_E:** ¿Cuán satisfecho esta Ud. con los siguientes aspectos?: Su empleo en esta IE
- **P1\_2:** EDAD
- **P1\_4:** En su hogar, ¿vive usted con personas de la tercera edad?
- **P1\_5:** En su hogar, ¿vive Ud. con personas que están en el grupo de riesgo ante COVID-19 por enfermedades preexistente

Durante el año 2020

- **P1\_11\_B:** ¿sufrió o sufre enfermedades respiratorias?
- **P1\_11\_F:** ¿sufrió o sufre ansiedad?
- **P1\_11\_G:** ¿sufrió o sufre depresión
- **P1\_11\_H:** ¿sufrió o sufre cancer?
- **P1\_11\_L:** ¿sufrió o sufre COVID-19?
- **P1\_18:** ¿En este momento se encuentra pagando algún préstamo o crédito?

```
data = endo%>%  
  select( P2_2, P1_24_E, P1_2, P1_4, P1_5, P1_11_B, P1_11_F, P1_11_G, P1_11_H, P1_11_L, P1_18)
```

Tenemos variable de sexo, edad, si es que es area rural o urbana. También si es que el docente vive con personas de tercera edad, o con personas que tienen factores de riesgo de COVID, si en el 2020 han tenido depresión, ansiedad, enfermedades respiratorias, también hay otra variable sobre si regresarían a clases de manera presencial.

Vamos a realizar diferentes modelos para calentar motores y volvernos expertas y expertos en la intepretación de coeficientes.

Lo que queremos hacer es ver **qué factores pueden influenciar en que un docente quiera retornar a clases presenciales**

**VARIABLE DEPENDIENTE: P2\_2 Retorno a clases**

```
table(data$P2_2)  
  
##  
##      0      1  
## 1551 16489  
  
data$retorno=as.factor(data$P2_2)  
levels(data$retorno) = c("No", "Si")  
table(data$retorno)  
  
##  
##    No    Si  
## 1551 16489
```

Ya teniendo lista la variable depediente vamos a realizar unos cuantos modelos y analizar el odds y la probabilidad.

## MODELO 1 : Retorno y personas de la tercera edad

- VD: Retorno
- VI: El docente vive con personas de la tercera edad P1\_4

```
data$P1_4=ifelse(data$P1_4 == "1", "1","0")
data$P1_4=as.numeric(data$P1_4) #numérica
table(data$P1_4)
```

```
##
##      0      1
## 11187  7772
```

Creemos nuestro modelo (función glm) Recuerda que lo que se está modelando es el logaritmo del odds (p/1-p).

```
modelo1 = glm(retorno ~ P1_4,family= binomial ,data)
summary(modelo1)
```

```
##
## Call:
## glm(formula = retorno ~ P1_4, family = binomial, data = data)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.53672    0.03715  68.287 < 2e-16 ***
## P1_4         -0.38607    0.05321  -7.255 4.01e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 10576  on 18039  degrees of freedom
## Residual deviance: 10524  on 18038  degrees of freedom
## (10176 observations deleted due to missingness)
## AIC: 10528
##
## Number of Fisher Scoring iterations: 5
```

Ahora vamos a calcular los coeficientes del modelo [los coeficientes obtenidos en esta regresión logística son el logaritmo natural de odds] [La función exponencial es la inversa del logaritmo]

Veamos

Revisemos los coeficientes.

```
coef(modelo1)
```

```
## (Intercept)      P1_4
##    2.536717   -0.386074
```

Recuerda que es importante revisar el signo del coeficiente, ya que dependiendo de eso procederemos a interpretar. En este caso, el coeficiente es negativo; es decir, la relación es inversa.



[Cuando calculamos el exponencial de los coeficientes, obtenemos el odds ratio (# de veces de que ocurra)]. Sin embargo, como el coeficiente es negativo esa exponencial debe dividir al 1 ( $1/\exp(\log)$ ).

```
1/exp(coef(modelo1))
```

```
## (Intercept)      P1_4
##  0.07912577  1.47119355
```

Siempre que la relación es inversa, la interpretación es cuando la VI disminuye, la VD aumenta en xx veces. En este caso, 1.47 veces.

Tenemos dos maneras de poner analizar este resultado Recordemos que la VI era si el o la docente vive con personas de la tercera edad.

- Si el o la docente vive con personas de la tercera edad, el odds de que quiera retornar a clases presenciales disminuye en 1.47
- Otra manera analizarlo como una probabilidad, para ello es necesario realizar el siguiente cálculo  $(1-\exp(\text{coef}))*100$

Recuerda que el coeficiente de la VI es  $-0.3860$  es el

```
(1-exp(-0.3860))*100
```

```
## [1] 32.02295
```

En este caso podemos interpretar los resultados como “si el docente vive con una persona de la tercera edad, la probabilidad de que quiera retornar a clases presenciales DISMINUYE en un 32.03%.

## MODELO 2

**REALICEMOS UN SEGUNDO MODELO** Agreguemos más variables:

- vive con personas de la tercera edad (P1\_4)
- ¿sufrió o sufre cancer? (P1\_11\_H)
- ¿sufrió o sufre depresión? (P1\_11\_G)

Queremos saber si estas variables influyen en la probabilidad de que el docente quiera retornar o no a clases presenciales

```
modelo2 = glm(retorno ~ P1_4+P1_11_H+P1_11_G, family = binomial,data = data)
summary(modelo2)
```

```
##
## Call:
## glm(formula = retorno ~ P1_4 + P1_11_H + P1_11_G, family = binomial,
##      data = data)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.59328     0.03944  65.760 < 2e-16 ***
## P1_4         -0.38429     0.05328  -7.213 5.46e-13 ***
## P1_11_H      -0.92276     0.21281  -4.336 1.45e-05 ***
## P1_11_G      -0.26229     0.06741  -3.891 9.99e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 10576  on 18039  degrees of freedom
## Residual deviance: 10493  on 18036  degrees of freedom
## (10176 observations deleted due to missingness)
## AIC: 10501
##
## Number of Fisher Scoring iterations: 5
```

```
coef(modelo2)
```

```
## (Intercept)      P1_4      P1_11_H      P1_11_G
##  2.5932761 -0.3842897 -0.9227575 -0.2622864
```

**Ojo, los tres coeficientes son negativos. Calculemos el exponencial**

Para calcular el odds ratio.

```
exp(coef(modelo2))
```

```
## (Intercept)      P1_4      P1_11_H      P1_11_G
## 13.3735135  0.6809341  0.3974216  0.7692907
```

Nuevamente, el odds ratio es menor que 1.

Como son menores que 1 entonces lo restamos y explicamos los resultados en base a la disminución de veces.

```
1/(exp(-0.38429))
```

```
## [1] 1.468571
```

```
1/(exp(-0.92276))
```

```
## [1] 2.516226
```

```
1/(exp(-0.26229))
```

```
## [1] 1.299903
```

Análisis según n° de veces

- Si el docente vive con personas de la tercera edad el odds/probabilidad de que quiera retornar a clases presenciales disminuye en 1.46 veces
- Si el docente ha tenido o tiene cáncer el odds/probabilidad de que quiera retornar a clases presenciales disminuye en 2.51 veces
- Si el docente ha tenido o tiene depresión el odds/probabilidad de que quiera retornar a clases presenciales disminuye en 1.29 veces

Ahora analicemos las probabilidades:

```
#Cuando el odds es menor a 1  
(1-(exp(-0.38429)))*100
```

```
## [1] 31.90661
```

```
(1-(exp(-0.92276)))*100
```

```
## [1] 60.25794
```

```
(1-(exp(-0.26229)))*100
```

```
## [1] 23.07121
```

- Si el docente vive con personas de la tercera edad, la probabilidad de que quiera retornar a clases presenciales disminuye en 31%
- Si el docente ha tenido o tiene cáncer la probabilidad de que quiera retornar a clases presenciales disminuye en 60%
- Si el docente ha tenido o tiene depresión la probabilidad de que quiera retornar a clases presenciales disminuye en 23%

**Si queremos calcular datos determinados** Ejemplo 1: Si el docente no vive con personas de la tercera edad, tiene cancer y tiene depresión

```
log.odds1 = predict(modelo2, data.frame(P1_4 = 0, P1_11_H = 1, P1_11_G = 1))  
log.odds1
```

```
##          1  
## 1.408232
```

```
exp(log.odds1)/(1+exp(log.odds1))
```

```
##          1  
## 0.803487
```

La probabilidad de que quiera retornara a clases presenciales es de 0.80

Ejemplo 2: Si el docente no vive con personas de la tercera edad, no tiene cancer y tiene depresión

```
log.odds1 = predict(modelo2, data.frame(P1_4 = 0, P1_11_H = 0, P1_11_G = 1))  
exp(log.odds1)/(1+exp(log.odds1)) #lo pasamos a probabilidades
```

```
##          1  
## 0.9114113
```

La probabilidad de que quiera retornara a clases presenciales es de 0.91

## Ayuda divina

Les proponemos esta función (Agradézcanle a su profesor) que facilita la interpretación de los resultados. La función se llama Divine.Help, para poder usarla solo necesitas indicar como argumento al nombre de tu modelo: Divine.Help(modelo). Recuerda que esta función solo podrá ejecutarse si previamente has ejecutado el código que crea la función.

```
Divine.Help <- function(model) {  
  # Extraer los coeficientes del modelo  
  coeficients <- coef(model)[-1] # Excluye el intercepto  
  # Inicializar un vector para almacenar los efectos  
  effects <- numeric(length(coeficients))  
  
  for (i in 1:length(coeficients)) {  
    if (coeficients[i] < 1) {  
      effects[i] <- round(1 - exp(coeficients[i]), 2)  
    } else {  
      effects[i] <- round(exp(coeficients[i] - 1), 2)  
    }  
  }  
  
  # Generar la interpretación en lenguaje natural  
  interpretation <- paste0("Un aumento de una unidad en la variable ", names(coeficients),  
                           " está asociado con un cambio de ",  
                           abs(effects)*100, "% en la probabilidad de éxito.")  
  # Devolver los coeficientes, efectos y la interpretación en un dataframe  
  result <- data.frame(Coefficient = coeficients,  
                      Exp = exp(coeficients),  
                      Probability = effects,  
                      Interpretation = interpretation)  
  return(result)  
}  
  
Divine.Help(modelo2)
```

```
##           Coefficient      Exp Probability
## P1_4      -0.3842897 0.6809341          0.32
## P1_11_H   -0.9227575 0.3974216          0.60
## P1_11_G   -0.2622864 0.7692907          0.23
##
## P1_4      Un aumento de una unidad en la variable P1_4 está asociado con un cambio de 32% en la prob
## P1_11_H   Un aumento de una unidad en la variable P1_11_H está asociado con un cambio de 60% en la prob
## P1_11_G   Un aumento de una unidad en la variable P1_11_G está asociado con un cambio de 23% en la prob
```

## MODELO 3:

Nuestras explicativas serán si la persona vive o no con personas de la tercera edad, tiene o ha tenido ansiedad y la variable edad.

```
modelo3 = glm(retorno ~ P1_4+P1_11_F+P1_2, family = binomial, data = data)
summary(modelo3)
```

```
##
## Call:
## glm(formula = retorno ~ P1_4 + P1_11_F + P1_2, family = binomial,
##      data = data)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.547717   0.137744  25.756 < 2e-16 ***
## P1_4         -0.354417   0.053486  -6.626 3.44e-11 ***
## P1_11_F      -0.437741   0.055822  -7.842 4.44e-15 ***
## P1_2         -0.019287   0.002844  -6.782 1.19e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 10575  on 18034  degrees of freedom
## Residual deviance: 10414  on 18031  degrees of freedom
## (10181 observations deleted due to missingness)
## AIC: 10422
##
## Number of Fisher Scoring iterations: 5
```

Recuerda revisar los signos, para poder identificar el tipo de relación.

```
coef(modelo3)
```

```
## (Intercept)          P1_4          P1_11_F          P1_2
##  3.54771740 -0.35441669 -0.43774133 -0.01928669
```

Interpretemos según n° de veces



```
1/(exp(-0.35441669))
```

```
## [1] 1.425349
```

```
1/(exp(-0.43774133))
```

```
## [1] 1.549204
```

```
1/(exp(-0.01928669 ))
```

```
## [1] 1.019474
```

Análisis

- Si el docente vive con personas de la tercera edad el odds de que quiera retornar a clases presenciales disminuye en 1.42 veces
- Si el docente ha tenido o tiene ansiedad el odds de que quiera retornar a clases presenciales disminuye en 1.54 veces
- Si el docente aumenta en un 1 su edad el odds de que quiera retornar a clases presenciales disminuye en 1.01 veces

```
(1-(exp(-0.35441669))) *100
```

Ahora en porcentaje (probabilidad)

```
## [1] 29.84174
```

```
(1-(exp(-0.43774133)))*100
```

```
## [1] 35.45073
```

```
(1-(exp(-0.01928669 )))*100
```

```
## [1] 1.910189
```

- Si el docente vive con personas de la tercera edad la probabilidad de que quiera retornar a clases presenciales disminuye en 29%
- Si el docente ha tenido o tiene ansiedad la probabilidad de que quiera retornar a clases presenciales disminuye en 35.5%
- Si el docente aumenta en 1 su edad la probabilidad de que quiera retornar a clases presenciales disminuye en 1.9%

Ahora obtengamos la probabilidad según casos

```
log.odds3 = predict(modelo3, data.frame(P1_4 = 0, P1_11_F = 1, P1_2 = 50))  
exp(log.odds3)/(1+exp(log.odds3))#para pasarlo a probabilidad
```

```
##          1  
## 0.8952608
```

Cuando un o una docente vive con personas de tercera edad, tiene ansiedad y tenga 50 años, la probabilidad de que quiera retornar a clases presenciales es de 0.89.

Ahora obtengamos la probabilidad con menor edad.

```
log.odds3 = predict(modelo3, data.frame(P1_4 = 0, P1_11_F = 1, P1_2 = 25))  
exp(log.odds3)/(1+exp(log.odds3))
```

```
##          1  
## 0.93263
```

Cuando un o una docente vive con personas de tercera edad, tiene ansiedad y tenga 25 años, la probabilidad de que quiera retornar a clases presenciales es de 0.93.