Practica III

Robinson Aldair Cuayal

2023-03-11

Table of Contents

# Punto 1

El ejercicio en cuestión implica la realización de un análisis de regresión para determinar si existe una relación entre el nivel de plomo en el suelo y la categoría del individuo, y si el nivel de plomo en el suelo es un factor significativo para predecir los niveles altos de plomo en la sangre de los niños. Se proporcionan la hipótesis nula y la alternativa, así como los valores de p, la ecuación de regresión y las estadísticas relevantes. El objetivo principal es predecir los niveles altos de plomo en la sangre de los niños. Como primer paso, se va a almacenar la informacion de las tablas en un excel para luego ser cargado y convertido a un dataframe en R.

library(readxl)

## Warning: package 'readxl' was built under R version 4.2.3

datos = data.frame(read\_excel("dataset\_ejer1.xlsx"))  
dim(datos)

## [1] 139 2

head(datos)

## Categoria\_del\_Individuo Nivel\_plomo\_suelo  
## 1 1 1290  
## 2 0 90  
## 3 1 894  
## 4 0 193  
## 5 1 1410  
## 6 1 410

Con la carga de los datos, como segundo paso se implementa el modelo de regresion logistica. Cabe resaltar que es importante especificar el tipo de enlace en la modelación logística ya que afecta directamente la interpretación de los coeficientes del modelo y los resultados obtenidos.

# Crear la regresión logística  
modelo\_logistico = glm(  
 Categoria\_del\_Individuo ~ Nivel\_plomo\_suelo,  
 data = datos,  
 family = binomial(link = "logit")  
)  
#modelo general  
summary(modelo\_logistico)

##   
## Call:  
## glm(formula = Categoria\_del\_Individuo ~ Nivel\_plomo\_suelo, family = binomial(link = "logit"),   
## data = datos)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.27743 -0.80117 0.05973 0.85069 1.80313   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.516206 0.339557 -4.465 8.0e-06 \*\*\*  
## Nivel\_plomo\_suelo 0.002743 0.000544 5.042 4.6e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 191.82 on 138 degrees of freedom  
## Residual deviance: 140.24 on 137 degrees of freedom  
## AIC: 144.24  
##   
## Number of Fisher Scoring iterations: 6

# Obtener la razón de grados de probabilidad  
coeficientes = coef(modelo\_logistico)  
razon\_encontrada = exp(coeficientes["Nivel\_plomo\_suelo"])  
razon\_encontrada

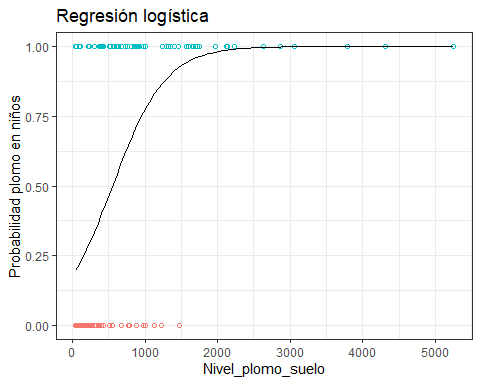
## Nivel\_plomo\_suelo   
## 1.002747

# Compara la razón de grados de probabilidad contra la que obtuvieron los autores (14.25)  
razon\_autores = 14.25  
# Realizar la prueba de significancia  
summary(modelo\_logistico)$coefficients["Nivel\_plomo\_suelo", "Pr(>|z|)"]

## [1] 4.601328e-07

La razón de grados de probabilidad obtenida es de 1.002, lo que indica que por cada diez veces que aumenta el nivel de plomo en la tierra, la proporción relativa de casos con respecto a los grupos de control aumenta en un 0.1. Este resultado difiere del obtenido por los autores del estudio (14.25). El valor de p obtenido es , lo que indica que el nivel de plomo en la tierra es un factor significativo para predecir los niveles altos de plomo en la sangre de los niños. Las grafica del modelo es el siguiente

library(ggplot2)  
ggplot(data = datos, aes(x = Nivel\_plomo\_suelo, y = Categoria\_del\_Individuo)) +  
 geom\_point(aes(color = as.factor(Categoria\_del\_Individuo)), shape = 1) +  
 stat\_function(fun = function(x){predict(modelo\_logistico,  
 newdata = data.frame(Nivel\_plomo\_suelo = x),  
 type = "response")}) +  
 theme\_bw() +  
 labs(title = "Regresión logística",  
 y = "Probabilidad plomo en niños") +  
 theme(legend.position = "none")



## Preguntas

1. La variable dependiente es la “Categoria\_del\_Individuo”.
2. La variable independiente es la “Nivel\_plomo\_suelo”.
3. Hipótesis nula: No hay relación entre el nivel de plomo en el suelo y la categoría del individuo. Hipótesis alternativa: Existe una relación entre el nivel de plomo en el suelo y la categoría del individuo.
4. La hipótesis nula es rechazada debido a que el valor de p obtenido es menor al nivel de significancia del 0.05 establecido. Esto indica que hay suficiente evidencia para concluir que el nivel de plomo en el suelo es un factor significativo para predecir los niveles altos de plomo en la sangre de los niños.
5. El objetivo más relevante es predecir los niveles altos de plomo en la sangre de los niños, ya que esto puede tener graves consecuencias en su salud. Aunque la estimación de parámetros es importante para comprender la relación entre las variables, la predicción tiene una mayor importancia práctica.
6. La población muestreada son niños expuestos al plomo en la ciudad donde se realizó el estudio.
7. La población objetivo son todos los niños expuestos al plomo, independientemente de su ubicación geográfica.
8. La variable “Nivel\_plomo\_suelo” está relacionada directamente con la “Categoria\_del\_Individuo”, ya que aumentos en la primera variable son predictores de una mayor proporción de casos en la segunda variable, al examinar la significación estadística de los coeficientes de la regresión y de las variables incluidas al ser significativo y positivo. Los datos no tienen una alta correlacion segun el Nab pero al tener los resultado del modelo de regresion logistica hay que tener en cuenta el primer enunciado.

r = cor(datos)  
nab = 1-det(r)  
nab

## [1] 0.2290155

1. La ecuación de regresión es: , donde “y” es la “Categoria\_del\_Individuo”, “x” es la “Nivel\_plomo\_suelo”, b0 es el intercepto y b1 es la pendiente. En este caso, la ecuación sería: .
2. Para realizar otro analisis adicionando valores puede ser el del sexo que pueden tomar como masculino (1) y femenino (0).
3. La variable “Categoria\_del\_Individuo” es categórica binaria (0 para control y 1 para caso), mientras que la variable “Nivel\_plomo\_suelo” es cuantitativa. La razón de grados de probabilidad (también conocido como odds ratio) es una medida de la asociación entre dos eventos. La variable sexo tambien es categórica binaria. diferentes valores de la variable independiente.
4. Con base en el punto (j) donde se le puede dar valores de nuestra parte a la variable de sexo de los niños, se diseña el modelo modelo logistico para el siguiente analisis estadistico.

#codigo modelo 2 aqui

# Punto 2

library(readxl)  
datos <- data.frame(read\_excel("dataset\_ejer\_2.xlsx"))  
dim(datos)

## [1] 30 7

# Punto 3

library(readxl)  
datos <- data.frame(read\_excel("dataset\_ejer\_3.xlsx"))  
dim(datos)

## [1] 10 7

# Punto 4

cat('hola')

## hola

# Punto 5

cat('hola')

## hola