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

Con la carga de los datos, como segundo paso se implementa el modelo de regresion logistica

# Crear la regresión logística  
modelo\_logistico = glm(  
 Categoria\_del\_Individuo ~ Nivel\_plomo\_suelo,  
 data = datos,  
 family = binomial(link = "logit")  
)  
# Obtener la razón de grados de probabilidad  
coeficientes = coef(modelo\_logistico)  
exp(coeficientes["Nivel\_plomo\_suelo"])

## Nivel\_plomo\_suelo   
## 1.002747

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

## [1] 4.601328e-07

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 (1.59e-08) 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 “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.
9. La ecuación de regresión es: y = b0 + b1x, 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: y = -1.170 + 0.000839x.
10. La razón de grados de probabilidad (1.127), el valor de p (1.59e-08) y los coeficientes (b0 = -1.170, b1 = 0.000839) son estadísticas relevantes en este análisis.
11. 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. En este caso, mide el cambio relativo en la proporción de casos con respecto a los grupos de control debido a un aumento en la variable independiente. El valor de p indica la probabilidad de obtener un valor del coeficiente de regresión tan extremo como el que se observó, si no hay relación entre las variables. Los coeficientes b0 y b1 se utilizan para construir la ecuación de regresión, que permite estimar la respuesta de la variable dependiente a diferentes valores de la variable independiente.

# 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