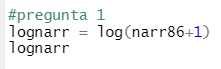
Regresión 2021.

Práctica de logística

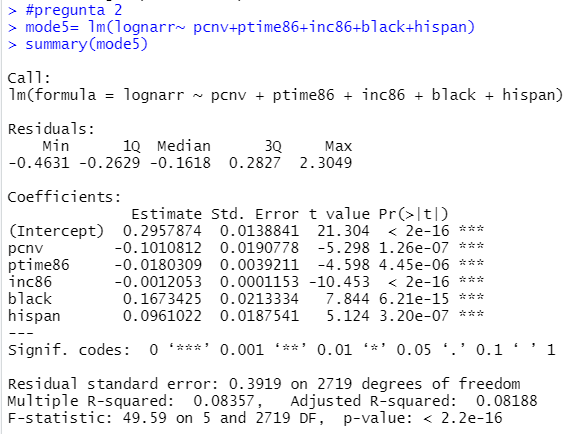
Se le da la base de datos crime1.Rdata. Esta contiene las siguientes variables:

|  |  |
| --- | --- |
| Variable | Explicación |
| narr86 | Número de arrestos |
| nfarr86 | Número de arrestos por contravención |
| nparr86 | Número de arrestos por crimen |
| pcnv | Proporción de declaratorias de culpabilidad previas al arresto |
| avgsen | Meses sentenciado a prisión |
| tottime | Meses en prisión desde los 18 años |
| ptime86 | Meses en prisión en el último año |
| qemp86 | Número de trimestres empleado |
| inc86 | Ingreso en cientos de dólares |
| durat | Duración del último período de desempleo |
| black | Dicotómica, 1=negro |
| hispan | Dicotómica, 1=hispano |
| born60 | Dicotómica, 1= nacido en 1960 |
| Pcnvsq | Cuadrado de pcnv |
| pt86sq | Cuadrado de ptime |
| inc86sq | Cuadrado de ingreso |

1. Cree una variable que se llame lognarr que sea igual al logaritmo natural de (narr86+1).



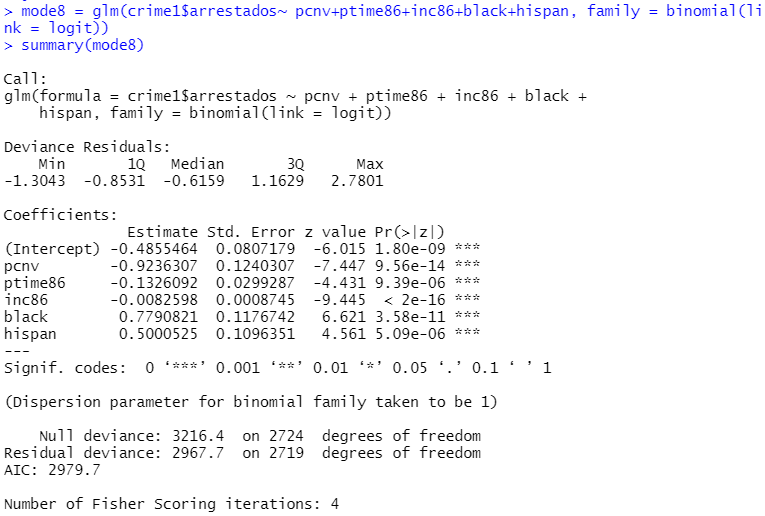
1. Estime un modelo gaussiano en el que lognarr esté en función de pcnv, ptime86, inc86, black y hispan



1. Con la función aic(), cálculele el AIC



1. Estime un modelo logístico en el que la variable arrestado sea predicha por pcnv, ptime86, inc86, black y hispan.



1. Llene la siguiente tabla

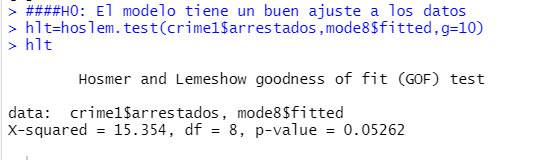
Cuadro 1. Resultados del modelo de regresión logística de condición de arrestado y regresión gaussiana del logaritmo del número de arrestos en función de la proporción de declaratorias de culpabilidad, meses en prisión, ingreso, condición de afroamericano y condición de hispano. A la par de cada coeficiente en la columna respectiva, coloque un asterisco si el p-value es menor a 0.05.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable |  | Gaussiano |  | Logística |  |
|  | Exp(coef) | sig |  | OR | sig |
| pcnv | 0.9038596 | \* |  | 0.3970748 | \* |
| ptime86 | 0.9821307 | \* |  | 0.8758073 | \* |
| inc86 | 0.9987954 | \* |  | 0.9917742 | \* |
| black | 1.1821591 | \* |  | 2.1794709 | \* |
| hispan | 1.1008716 | \* |  | 1.6488079 | \* |
|  |  |  |  |  |  |
| AIC | 2636,089 |  |  | 2979,696 |  |
|  |  |  |  |  |  |

1. Compare ambos modelos. Llega a conclusiones distintas con la regresión gaussiana y con la regresión logística?

En ambos casos se logra apreciar que todas las variables son significativas

1. Interprete los Odds Ratio para inc86 e hispan.
2. Pruebe la bondad de ajuste del modelo con una prueba de Hosmer y Lemeshow, al 5%.



No hay suficiente evidencia estadística para rechazar la hipotesis nula de que el modelo tiene un buen ajuste de los datos

1. Analice la capacidad predictiva del modelo con una tabla de clasificación.
2. Haga un gráfico de residuos de deviancia vs. leverage, identificados por número de arrestos.

##11. Haga un listado de las características de los casos que tengan un leverage extremo o un residuo extremo (Ud. Decide el límite), y describa a esos casos.