Regresion Logistica 2021 Reto

Andrés Villarreal González

2023-10-19

Regresión Lógistica Multinomial

Librerías necesarias y lectura de datos

```
library(nnet)
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.3     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.3     ✓ tibble    3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr     1.3.0
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(caret)
```

```
## Loading required package: lattice
##
## Attaching package: 'caret'
##
## The following object is masked from 'package:purrr':
##
##     lift
```

```
library(MASS)
```

```
##
## Attaching package: 'MASS'
##
## The following object is masked from 'package:dplyr':
##
##     select
```

```r
data <- read.csv("data_2021_RegLog.csv")

head(data)
```

```
##     PRS RH    SR  TOUT  WSR WDR O3Concentracion
## 1 698.9 43 0.059 20.83  8.7 109               3
## 2 706.8 29 0.881 36.61 12.9  56               3
## 3 704.4 16 0.946 34.57 10.0  64               3
## 4 706.8 49 0.336 30.92 10.0 136               3
## 5 697.5 37 0.041 28.39 14.1 102               3
## 6 697.8 46 0.044 26.42 11.7 107               3
```

Dividir datos en entrenamiento y prueba

```r
set.seed(123)

trainIndex <- createDataPartition(data$O3Concentracion, p = 0.8, list = FALSE)

train_data <- data[trainIndex, ]
test_data <- data[-trainIndex, ]
```

Ajustar el modelo de regresión y Resumen

```r
set.seed(123)

# Ajustar el modelo
model <- multinom(O3Concentracion ~ ., data = train_data)
```

```
## # weights:  24 (14 variable)
## initial  value 23031.308020
## iter  10 value 5959.477674
## iter  20 value 4009.868276
## iter  30 value 3791.876428
## iter  40 value 3771.061101
## final  value 3734.906691
## converged
```

```r
summary(model)
```

```
## Call:
## multinom(formula = O3Concentracion ~ ., data = train_data)
##
## Coefficients:
##   (Intercept)          PRS          RH       SR       TOUT          WSR
## 2   -7.082130  0.006137756 -0.04822663 2.904467 0.07008665  0.01461195
## 3    6.148726 -0.024837003 -0.02497572 3.686489 0.23219355 -0.04726438
##            WDR
## 2 -0.004860281
## 3 -0.001101421
##
## Std. Errors:
```

```
##    (Intercept)          PRS          RH          SR       TOUT
WSR
## 2 1.717799e-03 0.0002891926 0.002106992 0.114601215 0.00675806
0.007843537
## 3 4.953108e-05 0.0019725671 0.009904873 0.002953627 0.03917913
0.033425292
##            WDR
## 2 0.0004807038
## 3 0.0016679317
##
## Residual Deviance: 7469.813
## AIC: 7497.813
```

Precisión del modelo

```r
# Predecir en el conjunto de prueba
predicted_probs <- predict(model, newdata = test_data, type = "probs")
predicted_class <- predict(model, newdata = test_data, type = "class")


# Convertir tanto las predicciones como la variable real a factores con
niveles 1, 2, 3 y 4.
predicted_class <- factor(predicted_class, levels = c(1, 2, 3))
true_class <- factor(test_data$O3Concentracion, levels = c(1, 2, 3))

# Ahora crea la matriz de confusión
confusionMatrix(predicted_class, true_class)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    1    2    3
##          1 4870  279   15
##          2   34   40    3
##          3    0    0    0
##
## Overall Statistics
##
##                Accuracy : 0.9368
##                  95% CI : (0.9299, 0.9433)
##     No Information Rate : 0.9357
##     P-Value [Acc > NIR] : 0.3812
##
##                   Kappa : 0.1814
##
##  Mcnemar's Test P-Value : <2e-16
##
## Statistics by Class:
##
##                      Class: 1 Class: 2 Class: 3
## Sensitivity            0.9931 0.125392 0.000000
```

```
## Specificity            0.1276 0.992483 1.000000
## Pos Pred Value         0.9431 0.519481      NaN
## Neg Pred Value         0.5584 0.945972 0.996566
## Prevalence             0.9357 0.060866 0.003434
## Detection Rate         0.9292 0.007632 0.000000
## Detection Prevalence   0.9853 0.014692 0.000000
## Balanced Accuracy      0.5603 0.558937 0.500000
```

Valor p de coeficientes

```
coefficients <- coef(model)
std_errors <- summary(model)$standard.errors
t_statistics <- coefficients / std_errors
df <- nrow(train_data) - length(coefficients)
p_values <- 2 * (1 - pt(abs(t_statistics), df))
p_values
```

```
##   (Intercept) PRS         RH SR        TOUT        WSR       WDR
## 2           0   0 0.00000000  0 0.00000e+00 0.06248631 0.0000000
## 3           0   0 0.01169095  0 3.14379e-09 0.15736783 0.5090357
```

Coeficientes del modelo

```
coef(model)
```

```
##   (Intercept)          PRS          RH       SR       TOUT          WSR
## 2   -7.082130  0.006137756 -0.04822663 2.904467 0.07008665  0.01461195
## 3    6.148726 -0.024837003 -0.02497572 3.686489 0.23219355 -0.04726438
##            WDR
## 2 -0.004860281
## 3 -0.001101421
```

Significancia de variables mediante la prueba de Wald

```
library(nnet)
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```
Anova(model, type="II")
```

```
## Analysis of Deviance Table (Type II tests)
##
```

```
## Response: O3Concentracion
##      LR Chisq Df Pr(>Chisq)
## PRS      9.95  2   0.006926 **
## RH     538.31  2  < 2.2e-16 ***
## SR     661.05  2  < 2.2e-16 ***
## TOUT   157.04  2  < 2.2e-16 ***
## WSR      5.56  2   0.062006 .
## WDR    103.88  2  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
library(car)
Anova(model, test="Wald")
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: O3Concentracion
##      LR Chisq Df Pr(>Chisq)
## PRS      9.95  2   0.006926 **
## RH     538.31  2  < 2.2e-16 ***
## SR     661.05  2  < 2.2e-16 ***
## TOUT   157.04  2  < 2.2e-16 ***
## WSR      5.56  2   0.062006 .
## WDR    103.88  2  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```