

▾ A5-Regresión logística

Jorge Eduardo de Leon Reyna - A00829759

```
1 # Leyendo librerias
2 install.packages("ISLR")
3 install.packages("tidyverse")
4 install.packages("VCD")
5 install.packages("mosaic")

Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)

Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)

Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)

Warning message:
"package 'VCD' is not available for this version of R

A version of this package for your version of R might be available elsewhere,
see the ideas at
https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages"
Warning message:
"Perhaps you meant 'vcd' ?"
Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)

1 library(ISLR)
2 library(tidyverse)
3 library(mosaic)

Registered S3 method overwritten by 'mosaic':
  method      from
fortify.SpatialPolygonsDataFrame ggplot2

The 'mosaic' package masks several functions from core packages in order to add
additional features. The original behavior of these functions should not be affected by this.

Attaching package: 'mosaic'

The following object is masked from 'package:Matrix':

  mean

The following objects are masked from 'package:dplyr':

  count, do, tally

The following object is masked from 'package:purrr':

  cross

The following object is masked from 'package:ggplot2':

  stat

The following objects are masked from 'package:stats':

  binom.test, cor, cor.test, cov, fivenum, IQR, median, prop.test,
  quantile, sd, t.test, var

The following objects are masked from 'package:base':

  max, mean, min, prod, range, sample, sum
```

```
1 # Analisis de datos
2 head(Weekly)
3 glimpse(Weekly)
4 summary(Weekly)
5 pairs(Weekly)
6 cor(Weekly[, -9])
7 attach(Weekly)
8 plot(Volume)
```

	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	1990	0.816	1.572	-3.936	-0.229	-3.484	0.1549760	-0.270	0
2	1990	-0.270	0.816	1.572	-3.936	-0.229	0.1485740	-2.576	0
3	1990	-2.576	-0.270	0.816	1.572	-3.936	0.1598375	3.514	1
4	1990	3.514	-2.576	-0.270	0.816	1.572	0.1616300	0.712	1
5	1990	0.712	3.514	-2.576	-0.270	0.816	0.1537280	1.178	1
6	1990	1.178	0.712	3.514	-2.576	-0.270	0.1544440	-1.372	0

Rows: 1,089

Columns: 9

```

$ Year      <dbl> 1990, 1990, 1990, 1990, 1990, 1990, 1990, 1990, 1990, ...
$ Lag1      <dbl> 0.816, -0.270, -2.576, 3.514, 0.712, 1.178, -1.372, 0.807, 0...
$ Lag2      <dbl> 1.572, 0.816, -0.270, -2.576, 3.514, 0.712, 1.178, -1.372, 0...
$ Lag3      <dbl> -3.936, 1.572, 0.816, -0.270, -2.576, 3.514, 0.712, 1.178, -...
$ Lag4      <dbl> -0.229, -3.936, 1.572, 0.816, -0.270, -2.576, 3.514, 0.712, ...
$ Lag5      <dbl> -3.484, -0.229, -3.936, 1.572, 0.816, -0.270, -2.576, 3.514,...
$ Volume     <dbl> 0.1549760, 0.1485740, 0.1598375, 0.1616300, 0.1537280, 0.154...
$ Today      <dbl> -0.270, -2.576, 3.514, 0.712, 1.178, -1.372, 0.807, 0.041, 1...
$ Direction  <dbl> 0, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, ...

```

Year	Lag1	Lag2	Lag3
Min. :1990	Min. :-18.1950	Min. :-18.1950	Min. :-18.1950
1st Qu.:1995	1st Qu.: -1.1540	1st Qu.: -1.1540	1st Qu.: -1.1580
Median :2000	Median : 0.2410	Median : 0.2410	Median : 0.2410
Mean :2000	Mean : 0.1506	Mean : 0.1511	Mean : 0.1472
3rd Qu.:2005	3rd Qu.: 1.4050	3rd Qu.: 1.4090	3rd Qu.: 1.4090
Max. :2010	Max. : 12.0260	Max. : 12.0260	Max. : 12.0260

Lag4	Lag5	Volume	Today
Min. :-18.1950	Min. :-18.1950	Min. :0.08747	Min. :-18.1950
1st Qu.: -1.1580	1st Qu.: -1.1660	1st Qu.:0.33202	1st Qu.: -1.1540
Median : 0.2380	Median : 0.2340	Median :1.00268	Median : 0.2410
Mean : 0.1458	Mean : 0.1399	Mean :1.57462	Mean : 0.1499
3rd Qu.: 1.4090	3rd Qu.: 1.4050	3rd Qu.:2.05373	3rd Qu.: 1.4050
Max. : 12.0260	Max. : 12.0260	Max. :9.32821	Max. : 12.0260

Direction
Min. :0.0000
1st Qu.:0.0000
Median :1.0000
Mean :0.5556
3rd Qu.:1.0000
Max. :1.0000

A matrix: 8 × 8 of type dbl

Year	Lag1	Lag2	Lag3	Lag4	Lag5	Volume	Today
------	------	------	------	------	------	--------	-------

```

1 # Calculo de modelo logístico
2 modelo.log.m <- glm(Direction ~ . -Today, data = Weekly, family = binomial)
3 summary(modelo.log.m)
4 contrasts(Direction)
5 confint(object = modelo.log.m, level = 0.95)
6

```

```

Call:
glm(formula = Direction ~ . - Today, family = binomial, data = Weekly)

1 # Modelo logístico con variables significativas
2 ggplot(data = Weekly, mapping = aes(x = Direction, y = Lag2)) +
3   geom_boxplot(aes(color = Direction)) +
4   geom_point(aes(color = Direction)) +
5   theme_bw() +
6   theme(legend.position = "null")
7

Lag2 | 0.015722 0.020407 0.031020 0.033077

1 # Training: observaciones desde 1990 hasta 2008
2 datos.entrenamiento <- (Year < 2009)

5 | 0.015722 0.020407 0.031020 0.033077

1 # Test: observaciones de 2009 y 2010
2 datos.test <- Weekly[!datos.entrenamiento, ]

1 # Verifica:
2 nrow(datos.entrenamiento) + nrow(datos.test)

0 200 400 600 800 1000

1 # Ajuste del modelo logístico con variables significativas
2 modelo.log.s <- glm(Direction ~ Lag2, data = Weekly, family = binomial, subset = datos.entrenamiento)
3 summary(modelo.log.s)

Call:
glm(formula = Direction ~ Lag2, family = binomial, data = Weekly,
    subset = datos.entrenamiento)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.20326    0.06428   3.162  0.00157 **
Lag2         0.05810    0.02870   2.024  0.04298 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1354.7 on 984 degrees of freedom
Residual deviance: 1350.5 on 983 degrees of freedom
AIC: 1354.5

Number of Fisher Scoring iterations: 4

1 # Vector con nuevos valores interpolados en el rango del predictor Lag2:
2 nuevos_puntos <- seq(from = min(Weekly$Lag2), to = max(Weekly$Lag2), by = 0.5)

1 # Predicción de los nuevos puntos según el modelo con el comando predict() se calcula la probabilidad de que la variable respue
2 predicciones <- predict(modelo.log.s, newdata = data.frame(Lag2 = nuevos_puntos), se.fit = TRUE, type = "response")

1 # Límites del intervalo de confianza (95%) de las predicciones
2 CI_inferior <- predicciones$fit - 1.96 * predicciones$se.fit
3 CI_superior <- predicciones$fit + 1.96 * predicciones$se.fit

1 # Matriz de datos con los nuevos puntos y sus predicciones
2 datos_curva <- data.frame(Lag2 = nuevos_puntos, probabilidad = predicciones$fit, CI.inferior = CI_inferior, CI.superior = CI_su

1 # Codificación 0,1 de la variable respuesta Direction
2 Weekly$Direction <- ifelse(Weekly$Direction == "Down", yes = 0, no = 1)

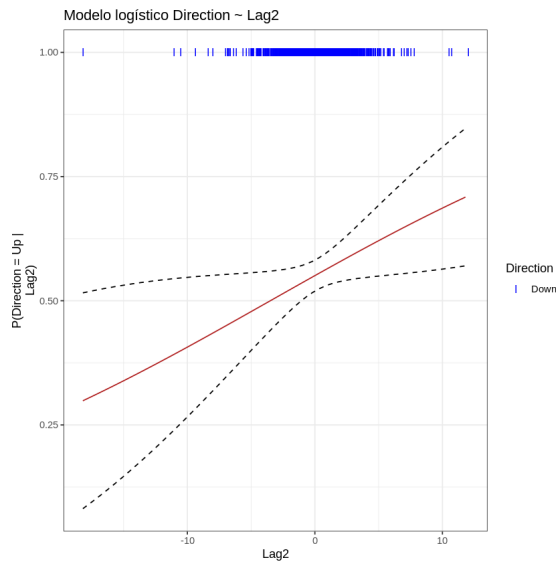
1 ggplot(Weekly, aes(x = Lag2, y = Direction)) +
2   geom_point(aes(color = as.factor(Direction)), shape = "I", size = 3) +
3   geom_line(data = datos_curva, aes(y = probabilidad), color = "firebrick") +
4   geom_line(data = datos_curva, aes(y = CI.superior), linetype = "dashed") +
5   geom_line(data = datos_curva, aes(y = CI.inferior), linetype = "dashed") +
6   labs(title = "Modelo logístico Direction ~ Lag2", y = "P(Direction = Up |
7   Lag2)", x = "Lag2") +
8   scale_color_manual(labels = c("Down", "Up"), values = c("blue", "red")) +

```

```

9 guides(color=guide_legend("Direction")) +
10 theme(plot.title = element_text(hjust = 0.5)) +
11 theme_bw()

```



```

1 # Chi cuadrada: Se evalúa la significancia del modelo con predictores con respecto al modelo nulo ("Residual deviance" vs "Null
2 anova(modelo.log.s, test = 'Chisq')

```

A anova: 2 × 5

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
	<int>	<dbl>	<int>	<dbl>	<dbl>
NULL	NA	NA	984	1354.710	NA
Lag2	1	4.166594	983	1350.543	0.04122861

```

1 # Cálculo de la probabilidad predicha por el modelo con los datos de test
2 prob.modelo <- predict(modelo.log.s, newdata = datos.test, type = "response")

```

```

1 # Vector de elementos "Down"
2 pred.modelo <- rep("Down", length(prob.modelo))
3 # Sustitución de "Down" por "Up" si la p > 0.5
4 pred.modelo[prob.modelo > 0.5] <- "Up"
5
6 Direction.0910 = Direction[!datos.entrenamiento]

```

```

1 # Matriz de confusión
2 matriz.confusion <- table(pred.modelo, Direction.0910)
3 matriz.confusion

```

```

      Direction.0910
pred.modelo 0 1
Down    9  5
Up     34 56

```

1