Instituto Tecnológico y de Estudios Superiores de Monterrey

A5-Regresión logística

Profesores:

Ivan Mauricio Amaya Contreras

Blanca Rosa Ruiz Hernandez

Antonio Carlos Bento

Frumencio Olivas Alvarez

Hugo Terashima Marín

Julian Lawrence Gil Soares – A00832272

19 de Octubre de 2023

```
In [ ]:   install.packages("IRkernel")
          IRkernel::installspec()
```

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

```
In [ ]:   install.packages('ISLR')
          library(ISLR)
          library(ggplot2)
          data("Weekly")
          df <- Weekly

          cor(Weekly[, -9])

          attach(Weekly)

          modelo.log.m <- glm(Direction ~ . -Today, data = Weekly, family = binomial)
          summary(modelo.log.m)
          contrasts(Direction)
          confint(object = modelo.log.m, level = 0.95)

          # Gráfico de las variables significativas (boxplot), ejemplo: Lag2):
          ggplot(data = Weekly, mapping = aes(x = Direction, y = Lag2)) +
          geom_boxplot(aes(color = Direction)) +
          geom_point(aes(color = Direction)) +
          theme_bw() +
          theme(legend.position = "null")

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

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

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

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

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

          predicciones <- predict(modelo.log.s, newdata = data.frame(Lag2 =nuevos_puntos),se.

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

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

```
# Codificación 0,1 de la variable respuesta Direction
Weekly$Direction <- ifelse(Weekly$Direction == "Down", yes = 0, no = 1)
ggplot(Weekly, aes(x = Lag2, y = Direction)) +
geom_point(aes(color = as.factor(Direction)), shape = "I", size = 3) +
geom_line(data = datos_curva, aes(y = probabilidad), color = "firebrick") +
geom_line(data = datos_curva, aes(y = CI.superior), linetype = "dashed") +
geom_line(data = datos_curva, aes(y = CI.inferior), linetype = "dashed") +
labs(title = "Modelo logístico Direction ~ Lag2", y = "P(Direction = Up |
Lag2)", x = "Lag2") +
scale_color_manual(labels = c("Down", "Up"), values = c("blue", "red")) +
guides(color=guide_legend("Direction")) +
theme(plot.title = element_text(hjust = 0.5)) +
theme_bw()

# Chi cuadrada: Se evalúa la significancia del modelo con predictores con respecto

anova(modelo.log.s, test ='Chisq')

# Cálculo de la probabilidad predicha por el modelo con los datos de test
prob.modelo <- predict(modelo.log.s, newdata = datos.test, type = "response")
# Vector de elementos "Down"
pred.modelo <- rep("Down", length(prob.modelo))
# Sustitución de "Down" por "Up" si la p > 0.5
pred.modelo[prob.modelo > 0.5] <- "Up"
Direction.0910 = Direction[!datos.entrenamiento]
# Matriz de confusión
matriz.confusion <- table(pred.modelo, Direction.0910)
matriz.confusión
library(vcd)
mosaic(matriz.confusion, shade = T, colorize = T,
gp = gpar(fill = matrix(c("green3", "red2", "red2", "green3"), 2, 2)))
mean(pred.modelo == Direction.0910)
```

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

A matrix: 8 × 8 of type dbl

	Year	Lag1	Lag2	Lag3	Lag4	Lag5	
Year	1.00000000	-0.032289274	-0.03339001	-0.03000649	-0.031127923	-0.030519101	0.8
Lag1	-0.03228927	1.000000000	-0.07485305	0.05863568	-0.071273876	-0.008183096	-0.0
Lag2	-0.03339001	-0.074853051	1.00000000	-0.07572091	0.058381535	-0.072499482	-0.0
Lag3	-0.03000649	0.058635682	-0.07572091	1.00000000	-0.075395865	0.060657175	-0.0
Lag4	-0.03112792	-0.071273876	0.05838153	-0.07539587	1.000000000	-0.075675027	-0.0
Lag5	-0.03051910	-0.008183096	-0.07249948	0.06065717	-0.075675027	1.000000000	-0.0
Volume	0.84194162	-0.064951313	-0.08551314	-0.06928771	-0.061074617	-0.058517414	1.0
Today	-0.03245989	-0.075031842	0.05916672	-0.07124364	-0.007825873	0.011012698	-0.0

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

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) 17.225822  37.890522   0.455   0.6494
Year        -0.008500   0.018991  -0.448   0.6545
Lag1        -0.040688   0.026447  -1.538   0.1239
Lag2         0.059449   0.026970   2.204   0.0275 *
Lag3        -0.015478   0.026703  -0.580   0.5622
Lag4        -0.027316   0.026485  -1.031   0.3024
Lag5        -0.014022   0.026409  -0.531   0.5955
Volume       0.003256   0.068836   0.047   0.9623
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1496.2  on 1088  degrees of freedom
Residual deviance: 1486.2  on 1081  degrees of freedom
AIC: 1502.2

Number of Fisher Scoring iterations: 4
```

A matrix: 2 ×

1 of type dbl

	Up
Down	0
Up	1

A matrix: 8 × 2 of type dbl

	2.5 %	97.5 %
(Intercept)	-56.985558236	91.66680901
Year	-0.045809580	0.02869546
Lag1	-0.092972584	0.01093101
Lag2	0.007001418	0.11291264
Lag3	-0.068140141	0.03671410
Lag4	-0.079519582	0.02453326
Lag5	-0.066090145	0.03762099
Volume	-0.131576309	0.13884038

```
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
```
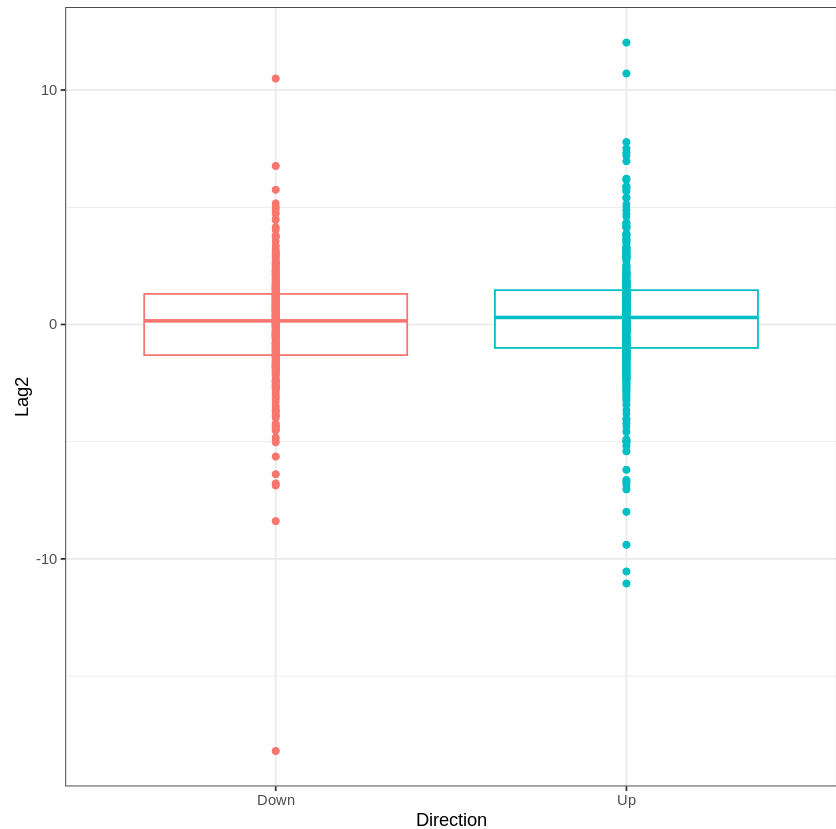
A anova: 2 × 5

	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

```
Error in eval(expr, envir, enclos): object 'matriz.confusión' not found
Traceback:
```

Modelo logístico Direction ~ Lag2