

A7-Regresión logística

Erika Martínez Meneses

2024-11-05

Trabaja con el set de datos Weekly, que forma parte de la librería ISLR. Este set de datos contiene información sobre el rendimiento porcentual semanal del índice bursátil S&P 500 entre los años 1990 y 2010. Se busca predecir el tendimiento (positivo o negativo) dependiendo del comportamiento previo de diversas variables de la bolsa bursátil S&P 500.

Encuentra un modelo logístico para encontrar el mejor conjunto de predictores que auxilien a clasificar la dirección de cada observación.

```
library(ISLR)
data <- Weekly
head(data)
```

##	Year	Lag1	Lag2	Lag3	Lag4	Lag5	Volume	Today	Direction
## 1	1990	0.816	1.572	-3.936	-0.229	-3.484	0.1549760	-0.270	Down
## 2	1990	-0.270	0.816	1.572	-3.936	-0.229	0.1485740	-2.576	Down
## 3	1990	-2.576	-0.270	0.816	1.572	-3.936	0.1598375	3.514	Up
## 4	1990	3.514	-2.576	-0.270	0.816	1.572	0.1616300	0.712	Up
## 5	1990	0.712	3.514	-2.576	-0.270	0.816	0.1537280	1.178	Up
## 6	1990	1.178	0.712	3.514	-2.576	-0.270	0.1544440	-1.372	Down

Se cuenta con un set de datos con 9 variables (8 numéricas y 1 categórica que será nuestra variable respuesta: Direction). Las variables Lag son los valores de mercado en semanas anteriores y el valor del día actual (Today). La variable volumen (Volume) se refiere al volumen de acciones. Today es el rendimiento de la semana actual. Y la variable Direction es una variable categórica que indica si el rendimiento es positivo ("Up") o negativo ("Down").

1. Análisis de datos.

```
library(tidyverse)
```

```
## — Attaching core tidyverse packages —————
tidyverse 2.0.0 —
## ✓ dplyr      1.1.4      ✓ readr      2.1.5
## ✓ forcats    1.0.0      ✓ stringr    1.5.1
## ✓ ggplot2     3.5.1      ✓ tibble     3.2.1
## ✓ lubridate  1.9.3      ✓ tidyr      1.3.1
## ✓ purrr      1.0.2
## — Conflicts —————
tidyverse_conflicts() —
```

```
## X dplyr::filter() masks stats::filter()
## X dplyr::lag() masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force
all conflicts to become errors
```

```
glimpse(Weekly)
```

```
## Rows: 1,089
## Columns: 9
## $ Year      <dbl> 1990, 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 <fct> Down, Down, Up, Up, Up, Down, Up, Up, Up, Down,
Down, Up, Up...
```

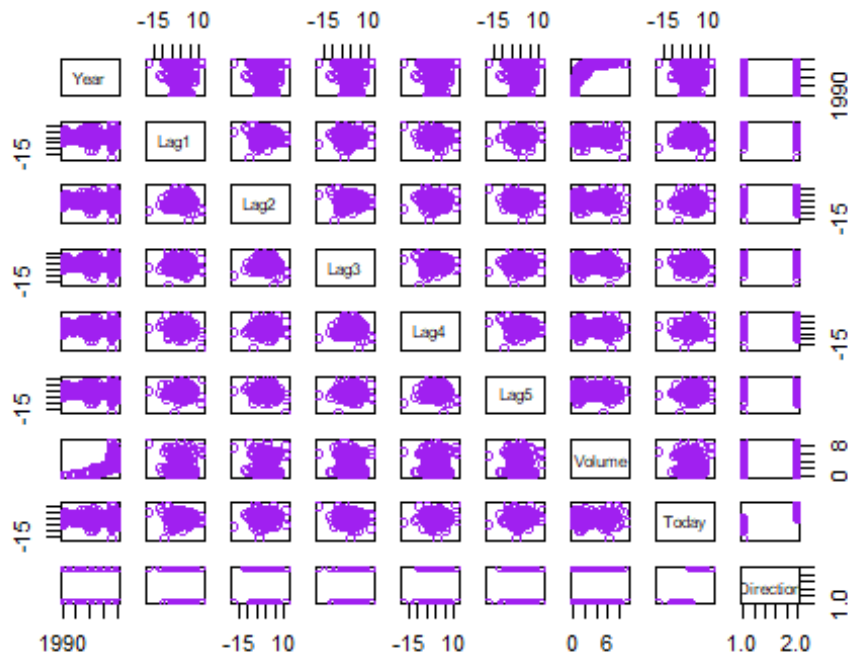
```
summary(data)
```

```
##      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
```

```
## Down:484
## Up :605
##
##
##
##
```

Las estadísticas descriptivas de summary nos permiten observar los valores promedios y dispersión de las variables. Los datos cubren el periodo de 1990 a 2010, con 1,089 observaciones. Las variables de Lag1 a Lag5 y Today muestran valores que varían entre -18.1950 y 12.0260 concentrándose alrededor de valores cercanos a cero, lo cual sugiere que las fluctuaciones semanales en el mercado son moderadas. La variable Volume, con un promedio de 1.57 y un rango de 0.08747 a 9.32821, indica variaciones significativas en el volumen de operaciones a lo largo del tiempo. La variable Direction muestra una ligera tendencia al alza en el mercado, con 605 observaciones etiquetadas como Up y 484 como Down.

```
pairs(Weekly, col = "purple")
```



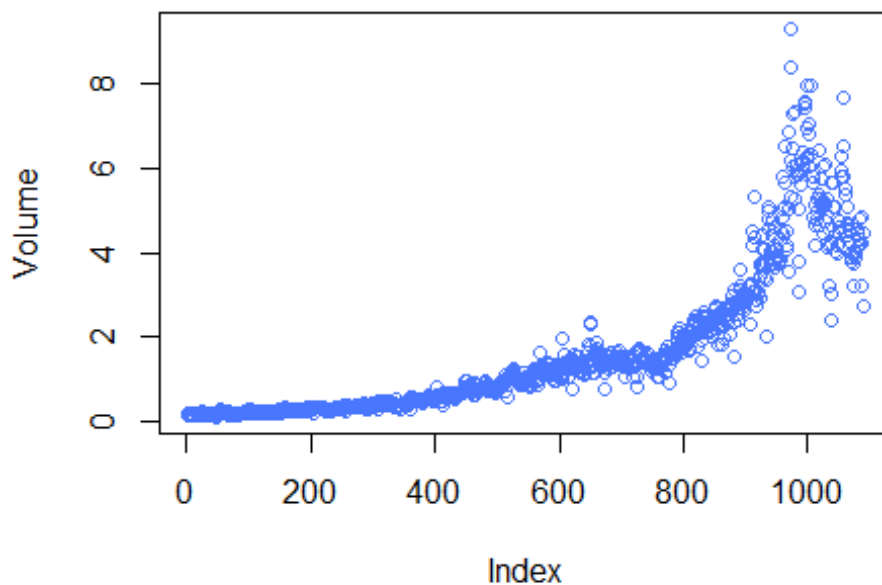
La gráfica anterior nos ayuda a visualizar la relación entre variables y lo que podemos interpretar es que no se observan correlaciones lineales fuertes entre las variables de rezago (Lag1 a Lag5) y la dirección del mercado (Direction), lo que indica que predecir Direction solo con estas variables podría ser difícil. Sin embargo, Volume muestra un aumento a lo largo del tiempo, lo que sugiere un incremento en la actividad del mercado en los años más recientes.

```
cor(Weekly[, -9])
```

```
##           Year           Lag1           Lag2           Lag3           Lag4
## Year      1.00000000 -0.032289274 -0.03339001 -0.03000649 -0.031127923
## Lag1     -0.03228927  1.000000000 -0.07485305  0.05863568 -0.071273876
## Lag2     -0.03339001 -0.074853051  1.00000000 -0.07572091  0.058381535
## Lag3     -0.03000649  0.058635682 -0.07572091  1.00000000 -0.075395865
## Lag4     -0.03112792 -0.071273876  0.05838153 -0.07539587  1.000000000
## Lag5     -0.03051910 -0.008183096 -0.07249948  0.06065717 -0.075675027
## Volume    0.84194162 -0.064951313 -0.08551314 -0.06928771 -0.061074617
## Today    -0.03245989 -0.075031842  0.05916672 -0.07124364 -0.007825873
##           Lag5           Volume           Today
## Year     -0.030519101  0.84194162 -0.032459894
## Lag1     -0.008183096 -0.06495131 -0.075031842
## Lag2     -0.072499482 -0.08551314  0.059166717
## Lag3      0.060657175 -0.06928771 -0.071243639
## Lag4     -0.075675027 -0.06107462 -0.007825873
## Lag5      1.000000000 -0.05851741  0.011012698
## Volume   -0.058517414  1.00000000 -0.033077783
## Today     0.011012698 -0.03307778  1.000000000
```

La matriz de correlación muestra que Year y Volume tienen una correlación positiva fuerte (0.84), indicando un aumento en el volumen de transacciones a lo largo del tiempo. Las correlaciones entre las variables Lag1 a Lag5 y Direction son bajas, lo que sugiere que los rendimientos de semanas previas tienen una relación débil con el rendimiento de la semana actual (Direction). Las demás correlaciones son débiles, con valores cercanos a cero, lo que sugiere poca relación lineal entre la mayoría de las variables.

```
attach(Weekly)
plot(Volume, col = "#4876FF")
```



La gráfica anterior muestra la variable Volume del conjunto de datos Weekly, representando el volumen de transacciones semanales. Se puede observar un crecimiento gradual en el volumen desde el inicio hasta alrededor del índice 800, seguido de un aumento acelerado y un comportamiento más variable hacia el final del período. Esto sugiere un incremento en la actividad de transacciones en el S&P 500 en los últimos años, lo que podría estar asociado a factores como un mayor interés en el mercado o cambios en las condiciones económicas.

2. Modelo Logístico con todas las variables

Se formula un modelo logístico con todas las variables menos la variable “Today”. Calcula los intervalos de confianza para las β_i . Detecta variables que influyen y no influyen en el modelo.

```
modelo.log.m <- glm(Direction ~ .-Today, data = Weekly, family =
binomial)
summary(modelo.log.m)

##
## 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

contrasts(Direction)

##      Up
## Down  0
## Up    1

confint(object = modelo.log.m, level = 0.95)

## Waiting for profiling to be done...

##              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

```

Este modelo utiliza todas las variables disponibles excepto Today, con el objetivo de predecir Direction. La salida del modelo sugiere que Lag2 es la única variable significativa al 5% ($p < 0.05$), mientras que las demás variables no tienen una influencia estadísticamente significativa en la predicción de Direction.

Un aumento en el rendimiento de la semana Lag2 incrementa ligeramente la probabilidad de que la dirección de esta semana sea “Up” (positiva). En términos de “odds” (momios), un aumento unitario en Lag2 multiplica las probabilidades por $e^{0.059} \approx 1.06$, lo cual representa un efecto modesto.

Los intervalos de confianza para Lag2 son estrechos, lo que sugiere cierta precisión en su estimación, pero las demás variables tienen intervalos amplios, indicando alta incertidumbre sobre su efecto.

3. Entrenamiento - Prueba

Divide la base de datos en un conjunto de entrenamiento (datos desde 1990 hasta 2008) y de prueba (2009 y 2010).

```
# Training: observaciones desde 1990 hasta 2008
train = data[data$Year < 2009,]
# Test: observaciones de 2009 y 2010
test = data[data$Year > 2008,]
# Verifica:
head(test)
```

##	Year	Lag1	Lag2	Lag3	Lag4	Lag5	Volume	Today	Direction	
##	986	2009	6.760	-1.698	0.926	0.418	-2.251	3.793110	-4.448	Down
##	987	2009	-4.448	6.760	-1.698	0.926	0.418	5.043904	-4.518	Down
##	988	2009	-4.518	-4.448	6.760	-1.698	0.926	5.948758	-2.137	Down
##	989	2009	-2.137	-4.518	-4.448	6.760	-1.698	6.129763	-0.730	Down
##	990	2009	-0.730	-2.137	-4.518	-4.448	6.760	5.602004	5.173	Up
##	991	2009	5.173	-0.730	-2.137	-4.518	-4.448	6.217632	-4.808	Down

4. Modelo Logístico con variables significativas

Se formula el modelo logístico sólo con las variables significativas en la base de entrenamiento.

```
# Ajuste del modelo Logístico con variables significativas
modelo.log.s <- glm(Direction ~ Lag2, data = train, family = binomial)
summary(modelo.log.s)
```

```
##
## Call:
## glm(formula = Direction ~ Lag2, family = binomial, data = train)
##
## 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
```

Este modelo se enfoca en Lag2, que ha demostrado tener un efecto positivo sobre Direction. El modelo simplificado ayuda a entender mejor la relación entre el rendimiento de hace dos semanas y la tendencia actual.

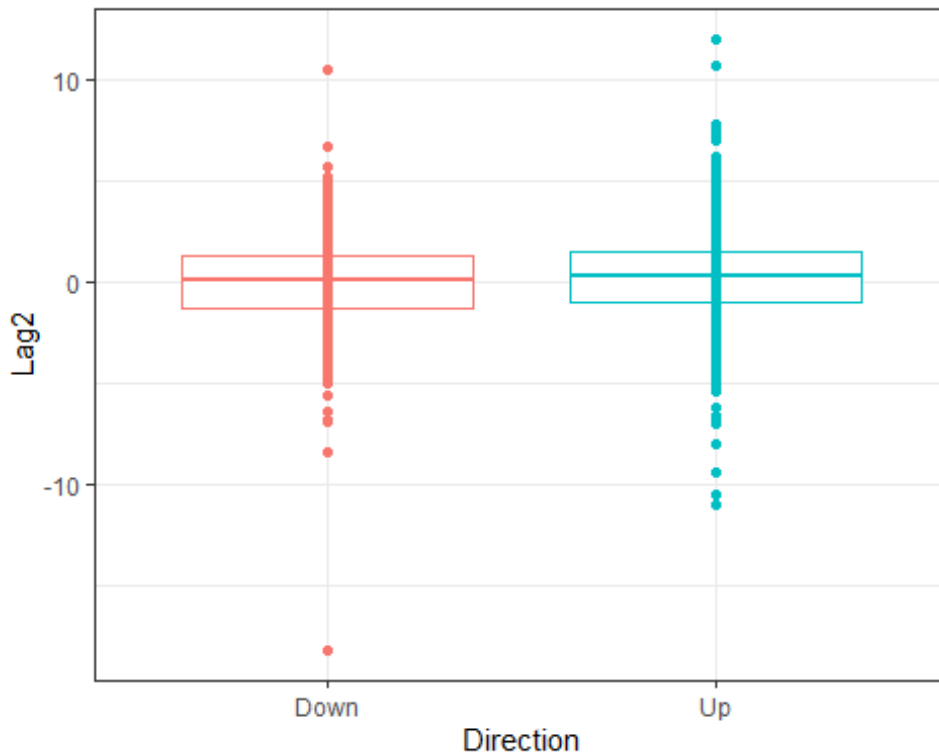
El modelo de regresión logística ajustado utiliza la variable Lag2 para predecir la dirección del mercado en la variable Direction y los resultados muestran que tanto el intercepto (0.20326) como el coeficiente de Lag2 (0.05810) son estadísticamente significativos, con valores de p de 0.00157 y 0.04298, respectivamente, lo que indica que la relación entre Lag2 y la dirección del mercado es significativa al nivel del 5%. Esto sugiere que un aumento en Lag2 está asociado a una mayor probabilidad de que la variable Direction sea "Up".

El modelo tiene un valor de devianza nula de 1354.7 y una devianza residual de 1350.5, lo que sugiere una ligera mejora en la capacidad del modelo para explicar la variabilidad de los datos al incluir Lag2. El AIC de 1354.5 proporciona una métrica de comparación para evaluar este modelo frente a otros posibles modelos. En general, aunque el modelo es significativo, la magnitud del coeficiente de Lag2 indica que el efecto es relativamente pequeño.

5. Representación Gráfica del Modelo

Representa gráficamente el modelo

```
# Visualización de Las Variables Significativas
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")
```

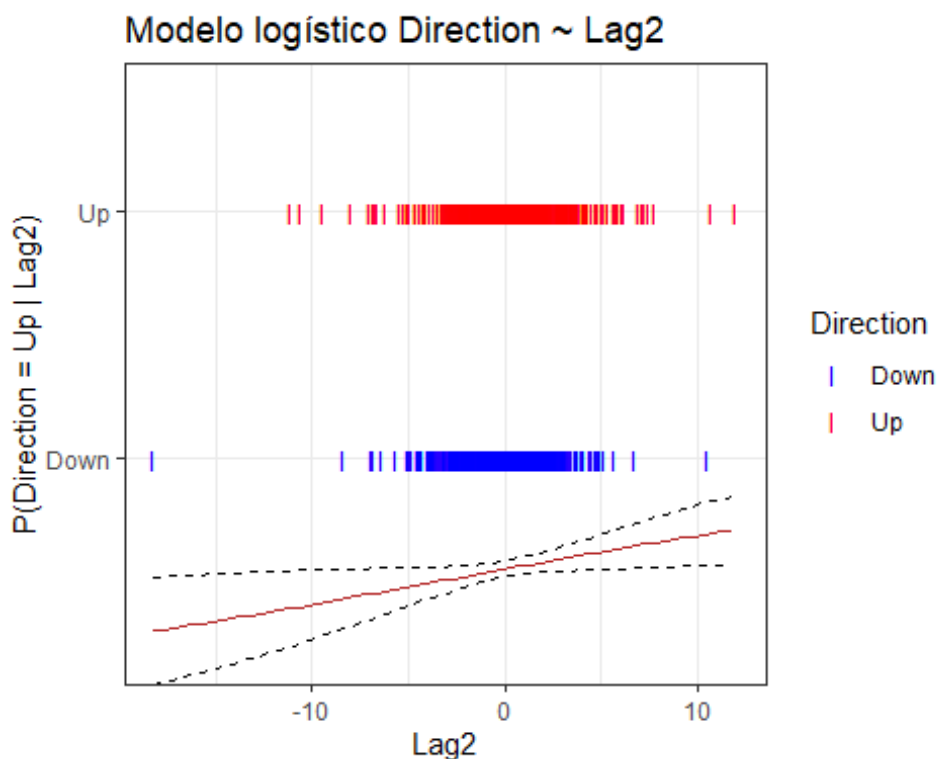
La gráfica muestra la distribución de la variable Lag2 en función de la dirección del mercado (Direction), que puede ser “Up” o “Down”. En general, la mediana de Lag2 para ambas direcciones está cerca de cero, aunque se observa una ligera diferencia en su dispersión. Para la dirección “Down”, la mayoría de los valores de Lag2 se agrupan en un rango más estrecho cerca de la mediana, mientras que para “Up” hay una mayor variabilidad con algunos valores extremos. Esta visualización respalda la idea de que Lag2 tiene una relación con la dirección del mercado, lo cual coincide con su significancia en el modelo logístico.

```
# Curva de Probabilidades del Modelo
# Generación de nuevos puntos y predicciones
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.fit = TRUE, type = "response")

# Cálculo de intervalos de confianza
CI_inferior <- predicciones$fit - 1.96 * predicciones$se.fit
CI_superior <- predicciones$fit + 1.96 * predicciones$se.fit

# Crear data para La gráfica
datos_curva <- data.frame(Lag2 = nuevos_puntos,
                          probabilidad = predicciones$fit,
                          CI.inferior = CI_inferior,
                          CI.superior = CI_superior)
```

```
# Gráfico de La curva logística con intervalos de confianza
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()
```

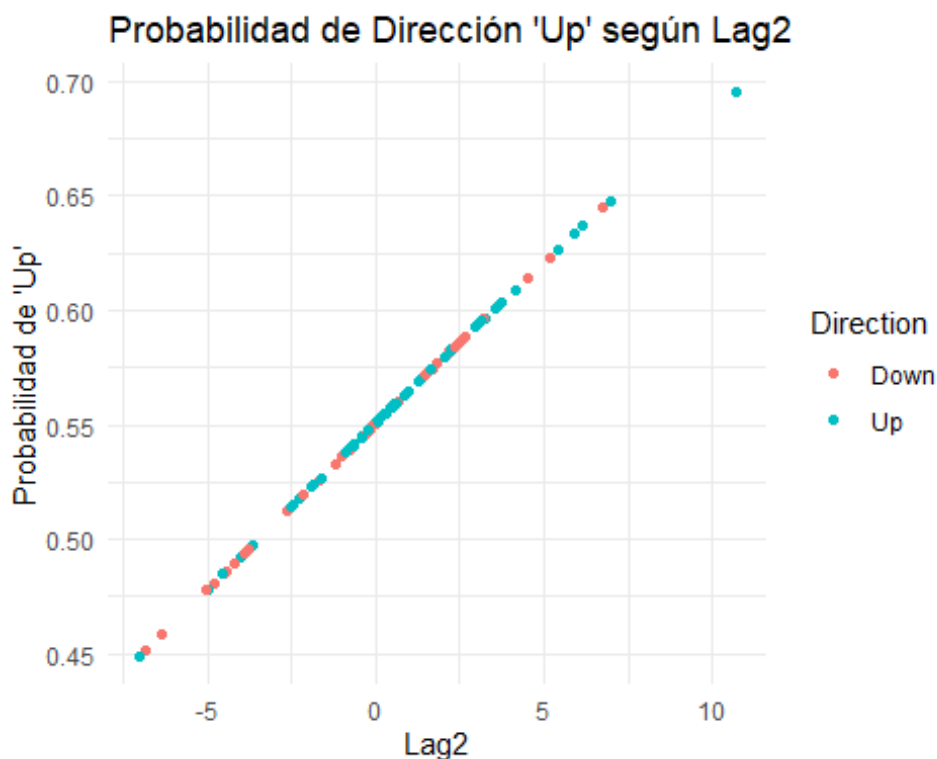


La gráfica anterior muestra el modelo de regresión logística ajustado para predecir la probabilidad de que la dirección del mercado (Direction) sea "Up" en función de la variable Lag2. Los puntos en color representan las observaciones individuales, con rojo para "Up" y azul para "Down". La línea continua en color rojo indica la probabilidad predicha de que Direction sea "Up" a medida que Lag2 aumenta, mientras que las líneas punteadas representan los intervalos de confianza para esta probabilidad.

Podemos observar que, a medida que el valor de Lag2 incrementa, la probabilidad de que la dirección sea “Up” también aumenta, aunque con una cierta dispersión, especialmente en valores extremos de Lag2. Esto sugiere una relación positiva entre Lag2 y la probabilidad de que el mercado suba, confirmando que Lag2 es un predictor relevante en el modelo.

```
# Asignar probabilidades y direcciones predichas al conjunto de prueba
test$pred_prob <- predict(modelo.log.s, newdata = test, type =
"response")
test$pred_direction <- ifelse(test$pred_prob > 0.5, "Up", "Down")

# Gráfica de probabilidad predicha en función de Lag2
ggplot(test, aes(x = Lag2, y = pred_prob, color = Direction)) +
  geom_point() +
  labs(title = "Probabilidad de Dirección 'Up' según Lag2",
       x = "Lag2", y = "Probabilidad de 'Up'") +
  theme_minimal()
```



Esta gráfica muestra la probabilidad predicha de que la dirección del mercado sea “Up” (alza) en función del valor de la variable Lag2 en el conjunto de prueba. Cada punto representa una observación con su probabilidad de “Up” predicha por el modelo logístico simple basado en Lag2, que fue ajustado en el conjunto de entrenamiento. Los puntos están coloreados según la dirección real observada (ya sea “Down” en rojo o “Up” en azul), permitiendo una comparación visual entre la predicción del modelo y la realidad.

A medida que el valor de Lag2 aumenta, la probabilidad de que la dirección sea “Up” también incrementa, lo cual indica una relación positiva entre Lag2 y la probabilidad de un movimiento al alza en el mercado. Esto sugiere que cuando Lag2 es más alto, el modelo asigna una mayor probabilidad de que el mercado suba en el siguiente periodo, y viceversa.

6. Evaluación del Modelo

Evalúa el modelo con las pruebas de verificación correspondientes (Prueba de chi cuadrada, matriz de confusión).

```
library(ggplot2)
library(vcd)

## Loading required package: grid

##
## Attaching package: 'vcd'

## The following object is masked from 'package:ISLR':
##
##      Hitters

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

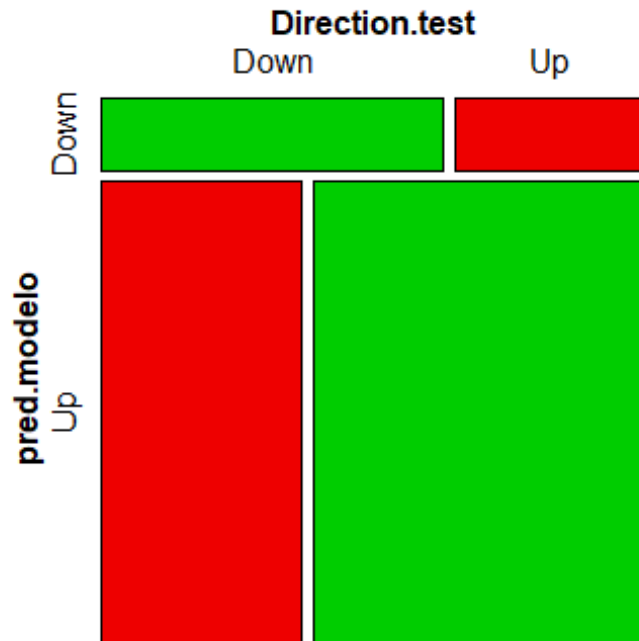
# Vector de elementos “Down” (predicción inicial)
pred.modelo <- rep("Down", length(prob.modelo))

# Sustitución de “Down” por “Up” si la probabilidad es mayor a 0.5
pred.modelo[pred.modelo > 0.5] <- "Up"

# Creación de la matriz de confusión
Direction.test <- test$Direction # Extrae la variable 'Direction' del conjunto de prueba
matriz.confusion <- table(pred.modelo, Direction.test)
matriz.confusion

##              Direction.test
## pred.modelo Down Up
##      Down      9  5
##      Up      34 56

# Visualización con un gráfico de mosaico
library(vcd)
mosaic(matriz.confusion, shade = TRUE, colorize = TRUE,
       gp = gpar(fill = matrix(c("green3", "red2", "red2", "green3"), 2,
2)))
```



```
# Cálculo de la precisión del modelo
precision <- mean(pred.modelo == Direction.test)
precision

## [1] 0.625
```

Para la evaluación del modelo se realizó una gráfica de mosaico y la matriz de confusión, estas muestran el desempeño del modelo logístico en la predicción de la dirección del mercado (“Up” o “Down”) en el conjunto de prueba. El modelo clasificó correctamente 9 casos como “Down” y 56 como “Up”, mientras que cometió errores en 5 casos donde predijo “Down” en lugar de “Up” y en 34 casos donde predijo “Up” en lugar de “Down”.

La precisión general del modelo, calculada como el porcentaje de predicciones correctas sobre el total, es aproximadamente 62.5%. Este valor sugiere que el modelo logra captar ciertos patrones en los datos, pero también presenta limitaciones, especialmente al sobreestimar las predicciones de “Up” cuando la dirección real es “Down”.

```
# Prueba de Chi-cuadrada
anova(modelo.log.s, test = 'Chisq')

## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Direction
```

```
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                984      1354.7
## Lag2  1    4.1666      983      1350.5  0.04123 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Asimismo, se realizó la prueba de ANOVA en el modelo logístico, esta muestra que la variable Lag2 tiene un efecto significativo en la predicción de la dirección del mercado (Direction). La reducción en la devianza (de 1354.7 a 1350.5) al incluir Lag2 en el modelo es estadísticamente significativa, con un valor p de 0.04123, lo cual es menor que 0.05. Esto indica que Lag2 contribuye de manera significativa a la capacidad predictiva del modelo, aunque su efecto no es muy fuerte.

7.

Escribe (ecuación), grafica el modelo significativo e interprétalo en el contexto del problema.

La ecuación para el modelo logístico que predice la probabilidad de que la variable Direction sea “Up” en función de Lag2 es $\text{logit}(p(\text{Direction} = \text{“Up”})) = \beta_0 + \beta_1 \cdot \text{Lag2}$ por lo que quedaría de la siguiente manera

```
cat("Ecuación del modelo logístico:\n")
## Ecuación del modelo logístico:
cat("logit(P(Direction = 'Up')) =", round(coef(modelo.log.s)[1], 4), "+",
    round(coef(modelo.log.s)[2], 4), "* Lag2\n\n")
## logit(P(Direction = 'Up')) = 0.2033 + 0.0581 * Lag2
```

Donde $P(\text{Direction} = \text{“Up”})$ es la probabilidad de que la dirección sea “Up” y β_0 y β_1 son los coeficientes que se estiman en el modelo.

En este contexto, el modelo logístico nos ayuda a entender cómo el valor de Lag2 (el rezago en el retorno del mercado de dos semanas atrás) afecta la probabilidad de que el mercado suba (“Up”) en la semana actual. Si β_1 es positivo y significativo, entonces un aumento en Lag2 está asociado con una mayor probabilidad de que Direction sea “Up”. Si es negativo, un aumento en Lag2 se asocia con una mayor probabilidad de que el mercado baje. El valor de β_1 también indica cuánto cambia la probabilidad de “Up” por cada unidad de cambio en Lag2. Si β_1 es cercano a cero, el efecto de Lag2 sobre la dirección del mercado es pequeño. En cambio, si es grande, Lag2 tiene un impacto considerable en la probabilidad de “Up”. Los intervalos de confianza nos muestran el rango dentro del cual cae la probabilidad predicha para un nivel de confianza del 95%. Esto proporciona una medida de la precisión del modelo. En nuestro caso el

coeficiente positivo de Lag2 (0.0581) indica una relación positiva entre esta variable y la probabilidad de “Up”. Así, cuando Lag2 aumenta, el modelo tiende a predecir una mayor probabilidad de que el mercado suba.

Aunque Lag2 es estadísticamente significativa, su efecto es modesto. Este modelo podría tener limitaciones al predecir la dirección semanal de la bolsa, ya que los cambios en el mercado pueden depender de más factores que no están presentes en este conjunto de datos. Por esto una posible mejora sería incorporar otras variables externas (como indicadores económicos) podría mejorar el modelo. Además, el modelo logístico asume una relación lineal entre Lag2 y la log-odds de Direction, lo cual podría no capturar la complejidad total del mercado.