



**CUNEF**

COLEGIO UNIVERSITARIO DE  
ESTUDIOS FINANCIEROS

## **PREDICCIÓN**

### **Práctica 1: Fondos**

V. Miguel Sempere Navarro

En la presente práctica de la asignatura de Predicción se propone analizar un dataset llamado 'Fondos' del cual se pueden extraer gran variedad de datos referentes a distintos fondos de inversión. Entre las diferentes variables que existen se encuentra la rentabilidad global del fondo, la gestora que lo comercializa, la volatilidad y rentabilidades en diferentes espacios temporales.

Puesto que, según el enunciado, la solución no es única, se ha optado por realizar un modelo de predicción de la rentabilidad global de los distintos fondos realizando para ello distintos análisis vistos durante la primera sesión; entre los que se encuentran distintas maneras de seleccionar variables.

El modelo empieza con la importación del dataset con el cual se va a trabajar:

```
##Práctica 1 de Predicción - FONDOS  
  
library(leaps)  
library(car)  
  
datosfondos <- read.csv('Fondos.csv', sep = ';', dec = ',')  
View(datosfondos)  
  
datosfiltrados <- datosfondos[, c(-1, -2, -5, -6, -18)]  
View(datosfiltrados)  
  
Y <- datosfondos$rent_1  
mX <- cbind(1, datosfondos[,10:16])
```

*Ilustración 1. Importación Dataset. Fuente: elaboración propia.*

Como se puede observar, se cargan dos librerías que nos harán falta en el futuro y lo más importante: se filtran los datos para eliminar variables que no aportan ningún tipo de valor cuantitativo al modelo (fechas, nombres de las gestoras, etc.). Se elige como variable independiente la variable Y que hace referencia a la rentabilidad global del fondo.

Se ha observado también que había algunas variables con un gran número de NAs, por lo que se ha optado por eliminarlas del modelo. Sin embargo, las variables con pocos NAs se dan por válidas, asumiendo que habrá un error adicional que no se puede remediar.

El siguiente paso realizado ha sido la realización de tres modelos de regresión lineal con el objetivo de escoger aquellas variables que más información aportan al modelo:

```
regres01 <- lm(Y ~ rent_1_mes + rent_3_meses + rent_6_meses + rent_en_el_año + Volatilidad_3, datosfondos)
summary(regres01) #Tras ver este summary, eliminamos algunas variables que tienen una Pr muy alta (menos estrellas)

regres02 <- lm(Y ~ ., datosfiltrados)
summary(regres02) #Tras ver este summary, eliminamos algunas variables que tienen una Pr muy alta (menos estrellas)

regres03 <- lm(Y ~ rent_6_meses + rent_en_el_año, datosfondos)
summary(regres03) #Tras ver este summary, eliminamos algunas variables que tienen una Pr muy alta (menos estrellas)
```

*Ilustración 2. Modelos de Regresión Lineal. Fuente: elaboración propia.*

Los dos primeros modelos han servido para realizar el tercero, pues a medida que se daban los resultados, según el PValor de las diferentes variables se escogían las mejores para realizar el mejor modelo. Sin embargo, ha sido necesario tener criterio, pues se ha observado que existen variables que, aunque su PValor era muy bajo, carecía de sentido incluirlas en el modelo de regresión porque la información no estaba relacionada; como por ejemplo el “estilo de inversión” (variable categórica).

Como modelo de regresión lineal definitivo se escoge el tercero, teniendo en cuenta las variables de la rentabilidad a 6 meses y a un año, porque son las que mejor explican la rentabilidad global.

Se han llevado a cabo un análisis de selección de variables AIC, un BIC y la prueba Best Subset:

```
#Modelos de selección de variables
AIC(regres03, regres02)
BIC(regres03, regres02)

#Modelo de selección Best Subset
regfit.full = regsubsets(Y~.,datosfiltrados )
reg.summary = summary(regfit.full)
head(reg.summary)

reg.summary$rss

reg.summary$cp

reg.summary$aic

reg.summary$bic
```

*Ilustración 3. Modelos de selección. Fuente: elaboración propia.*

El último paso ha sido realizar la predicción de la variable independiente (rentabilidad global) en base a las regresiones realizadas con anterioridad. Aunque se han realizado las predicciones con los tres modelos de regresión iniciales, se da por válida la última predicción (en base al tercer modelo de regresión lineal). Aquí se muestran los primeros valores y la comparación con los valores reales:

	V1	prediccion3
1	27.22	30.478927358
2	24.65	26.939557661
3	21.19	20.845152644
4	20.95	20.531229524
5	17.20	21.118119439
6	11.08	8.925632555
7	10.99	4.994729095
8	9.56	5.322989584
9	8.78	7.386453784
10	8.76	6.329887356
11	7.58	3.253603757
12	7.36	9.214682094
13	7.16	3.629710467
14	7.04	2.900687769
15	6.66	6.469604530
16	6.41	6.285844463
17	5.34	4.671487372
18	5.19	3.043553303
19	5.15	5.702633569
20	5.11	5.267820198
21	5.04	5.743964264
22	4.92	0.321322625
23	4.91	3.942573715
24	4.81	1.041010007
25	4.63	5.395568364
26	4.56	4.107553384
27	4.53	4.780625974

*Ilustración 4. Predicción de la rentabilidad. Fuente: elaboración propia.*

Se observa que la predicción tiene un error relativamente bajo, pero se preveía dicho error por culpa de los NAs.

En el documento anexo se puede ver todo el código para la realización del informe, en el que se explica con mayor detalle los resultados obtenidos.