Universidad Nacional Autónoma de México

Facultad de Ciencias

TAREA 6

Selección de modelos y regularización

Aguirre Armada Guillermo

Figueroa Torres Ivan Emiliano

Luna Gutiérrez Yanely

Ortiz Silva Ana Beatriz

PROFESOR DE ASIGNATURA:
Guillermina Eslava

PROFESOR DE ADJUNTO:
Sofía Guzman

18 de enero de 2021

CIUDAD UNIVERSITARIA, CD. MX.

Ejercicio 1.

Usamos la base `Boston` de la paquetería `MASS` para ajustar modelos de regresión. La base contiene 506 observaciones de 14 variables, 13 del tipo numéricas y una variable, `chas`, de tipo factor. Tomamos a la varible `crim` como nuestra variable predictora.

Realizamos una selección de modelos usando la paquetería `MASS` y leaps, para ajustar métodos: aditivo, step k=2, step k=log(506) y regsubset exhaustive; cómo se muestra en la siguiente tabla 1.

Utiliando Boxcox encontramos que con una $\lambda = 0.02$, la transformación óptima es la función *log()*.Los modelos obtenidos fueron:

- Aditivo: = log(crim) $\sim$.
- Step k=2:= log(crim) $\sim$ zn + indus + nox + age + rad + ptratio + black + lstat + medv.
- Step k=log(506)/regsubset:= log(crim)$\sim$ zn + nox + age + rad + black + lstat.

Al ajustar los modelos notamos que por el método step k=log(506) y por el regsubset exhaustive, llegabamos a la misma regresión con 6 variables explicativas.

	Aditivo	Step k=2	Step k=log(n)/Regsubsets
ECM Ap	0.58	0.58	0.596
ECM Val	0.623	0.599	0.607
p	13	9	6
R^2	0.875	0.875	0.872
$R^2 Adj$	0.872	0.873	0.871
AIC	1191.48	1184.10	1190.06
BIC	1254.88	1230.59	1223.87

Table 1: **Estadísticas e información adicional de los modelos no regularizados.** *Vemos que las estadísticas de los modelos no difieren en gran tamaño unas de otras, a pesar de contar con diferentes números de variables*

Haciendo una selección de modelos y aplicando métodos regularizados obtenemos el ajuste de 6 modelos distintos, para medir su alcance predictivo calculamos su tasa de error aparente y de prueba.

Modelos	Error aparente	Error prueba
R. Aditiva	58.1	62.3
Step k=2	58.2	59.9
Step k=log(n)/Regsubsets	59.6	60.7
Lasso	58.5	61.9
Ridge	60.4	63.2
Elasticnet	58.2	61.7

Table 2: **Tasas error en porcentajes de modelos ajustados.** *Las primeras tres tasas pruebas fueron calculadas por training/ test con 300 repeticones, las últimas tres con cv k=10.*

	TrnErr(Ap)	cv(k=10, B=1)	cvsd	betas(df)	lamda.min	lamda.lse
Lasso	0.587	0.608	0.033	10	0.021	0.124
Ridge	0.603	0.635	0.0347	14	0.184	0.425
Elasticnet	0.583	0.620	0.039	12	0.0183	0.206

Table 3: **Estadísticas e información adicional de modelos regularizados.** *Vemos que las cvsd son similares para todos los modelos pero difieren en las lamdas, siendo la más pequeña minima para elasticnet y la más pequeña lse para Lasso.*

Modelos	Inter	zn	indus	chas1	nox	rm	age	dis	rad	tax	ptratio	black	lstat	medv
R. Aditiva	-3.73	-0.01	0.02	-0.04	3.84	-0.04	0.00	-0.00	0.14	0.00	-0.04	-0.00	0.03	0.01
Step k=2	-4.10	-0.01	0.02		3.86		0.00		0.14		-0.04	-0.00	0.03	0.00
Step k=log(n)	-4.83	-0.01			4.71		0.01		0.14			0.00	0.03	
Lasso		-0.01	0.02		3.94		0.01	-0.02	0.14		-0.02	0.00	0.02	
Ridge		-0.01	0.01	0.02	3.46	-0.02	0.01	-0.04	0.10	0.00	-0.02	0.00	0.03	0.01
Elasticnet		-0.01	0.02		3.86	-0.01	0.01	-0.02	0.14		-0.03	0.00	0.03	0.00

Table 4: **Valor de los coeficientes** *asociados a los siete modelos seleccionados, recordemos que Step k=log(n) es el mismo modelo que Regsubsets.*

Podemos observar que a pesar de ajustar distintos modelos, con diferentes números de variables, gracias a la transformación log(), el poder predictivo de los modelos es similar en los 6 modelos, al igual que las estadísticas.
Si bien es con el modelo aditivo donde tenemos una tasa de error aparente menor (**58.1**),su tasa prueba es una de las más altas (**62.3**); caso contrario con el ajuste Step k=2 (**59.9**)que se mantiene baja en comparación de los demás modelos y cambia sólo un **0.7** por cierto.
Nosotros consideramos que el modelo que seleccionaríamos con respecto a su poder predictivo, ya que contempla a 9 de las 13 variables predictoras, tiene unas R^2ś que están por encima de las otras, aunque no sea muy significativo el cambio (0.001), es la que tiene menor AIC y la segunda que tiene menor BIC, incluyendo lo antes mencionado con respecto a las tasas de error.

Ejercicio 2.

En este ejercicio usamos la información de la base `riboflavin` del paquete `hdi`, la cual contiene 73 observaciones de 4088 variables predictoras (`x`) y una variable respuesta (`y`). Todas las variables son numéricas.
Ajustamos tres modelos regularizados usando la función `glmnet()` del paquete `glmnet`.

- *ridge* (`alpha = 0`)
- *lasso* (`alpha = 1`)
- *elastic net* (`alpha = 0.5`)

Para elegir el valor de λ óptimo utilizamos la función `cv.glmnet()` con k=10, la cual usa validación cruzada sobre 100 modelos con distintos valores de λ. Encontramos que el valor de λ que minimiza el error cuadrático medio para cada modelo, el cual se encuentra en la Tabla , así como el numero de coeficientes distintos de cero para dicho valor de λ.
En la Tabla se muestran los valores del error cuadrático medio aparente y validadas para cada modelo. Podemos observar que el ECM incrementa en promedio solo 0.20 en cada modelo y se mantiene bastante estable entre una y 100 repeticiones.

	ridge	*lasso*	*elastic net*
ECM aparente	0.028	0.042	0.037
ECM cv(k=10, B=1)	0.262	0.200	0.216
ECM cv(k=10, B=100)	0.262	0.210	0.220
λ_{minECM}	5.934	0.042	0.069

Table 5: **Tasas de error de los modelos regularizados y su correspondiente valor de λ_{min}**. Usamos 10 *folds* en validación cruzada, con una repetición y con 100 repeticiones.

Para los modelos que seleccionamos usando el criterio de λ_{min} calculamos el número de coeficientes que en valor absoluto son mayores 0.025:

ridge: 6		*lasso:* 29		*elastic net:* 42	
xLYSC_at	-0.033	xYOAB_at	-0.784	xYOAB_at	-0.563
xYEBC_at	-0.032	**xYEBC_at**	-0.531	**xYEBC_at**	-0.497
xYFIT_at	0.032	xLYSC_at	-0.270	xLYSC_at	-0.396
xSPOVAA_at	0.028	xSPOVAA_at	0.247	xSPOVAA_at	0.283
xYBFI_at	0.027	xYQJU_at	0.213	xYQJU_at	0.198
xHUTP_at	-0.027	xYXLD_at	0.197	xYBFI_at	0.173

Table 6: **Coeficientes con mayor valor absoluto** para cada uno de los modelos y su variable asociada.

	ridge	*lasso*	*elastic net*
λ_{minECM}	5.93	0.042	0.069
Num. betas (λ_{minECM})	4089	38	57
λ_{1se}	30.22	0.080	0.161
Num. betas (λ_{1se})	4089	29	42
λ_{maxdf}	-	0.006	0.012
Num. betas (λ_{maxdf})	-	72	90

Table 7: **Información adicional de los modelos.** Valores de λ que cumplen tres criterios (`lambda.min`, `lambda.1se`, `dfmax`) y su correspondiente número de coeficientes que son distintos de cero.

Anexo

Ejercicio 1

```
    Call: Modelo aditivo
lm(formula = log(crim) ~ ., data = Boston)

Residuals:
   Min     1Q Median     3Q    Max
-2.570 -0.555 -0.037  0.502  2.627

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.732873   0.868764   -4.30  2.1e-05 ***
zn          -0.011685   0.002250   -5.19  3.0e-07 ***
indus        0.019756   0.010016    1.97  0.04911 *
```

```
chas1       -0.048092   0.141711   -0.34  0.73448
nox          3.846813   0.633484    6.07  2.5e-09 ***
rm          -0.048985   0.073588   -0.67  0.50594
age          0.005988   0.002152    2.78  0.00561 **
dis         -0.005438   0.033840   -0.16  0.87241
rad          0.142856   0.010573   13.51  < 2e-16 ***
tax         -0.000132   0.000619   -0.21  0.83110
ptratio     -0.041137   0.022389   -1.84  0.06675 .
black       -0.001502   0.000441   -3.40  0.00072 ***
lstat        0.031705   0.009093    3.49  0.00053 ***
medv         0.010469   0.007267    1.44  0.15030
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.773 on 492 degrees of freedom
Multiple R-squared:  0.875,Adjusted R-squared:  0.872
F-statistic:  266 on 13 and 492 DF,  p-value: <2e-16

    Call: Step k=2
lm(formula = log(crim) ~ zn + indus + nox + age + rad + ptratio +
    black + lstat + medv, data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max
-2.5680 -0.5614 -0.0362  0.5004  2.6822

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.100828   0.642667   -6.38  4.0e-10 ***
zn          -0.012056   0.001963   -6.14  1.7e-09 ***
indus        0.019619   0.008676    2.26  0.02417 *
nox          3.866694   0.596785    6.48  2.2e-10 ***
age          0.005775   0.002016    2.86  0.00435 **
rad          0.140597   0.006007   23.41  < 2e-16 ***
ptratio     -0.040579   0.022274   -1.82  0.06908 .
black       -0.001463   0.000433   -3.38  0.00079 ***
lstat        0.033694   0.008596    3.92  0.00010 ***
medv         0.008872   0.006169    1.44  0.15099
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.771 on 496 degrees of freedom
Multiple R-squared:  0.875,Adjusted R-squared:  0.873
F-statistic:  387 on 9 and 496 DF,  p-value: <2e-16

    Call: Step k=log(n) o Regsubsets
lm(formula = log(crim) ~ zn + nox + age + rad + black + lstat,
    data = Boston)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-2.6807 -0.5935 -0.0218  0.4972  2.6319

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.833234   0.302201  -15.99  < 2e-16 ***
zn          -0.011106   0.001839   -6.04    3e-09 ***
nox          4.714298   0.507822    9.28  < 2e-16 ***
age          0.006579   0.001989    3.31  0.00101 **
rad          0.137756   0.005312   25.93  < 2e-16 ***
black       -0.001456   0.000433   -3.36  0.00084 ***
lstat        0.025020   0.006542    3.82  0.00015 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.777 on 499 degrees of freedom
Multiple R-squared:  0.872,Adjusted R-squared:  0.871
F-statistic:  568 on 6 and 499 DF,  p-value: <2e-16
```

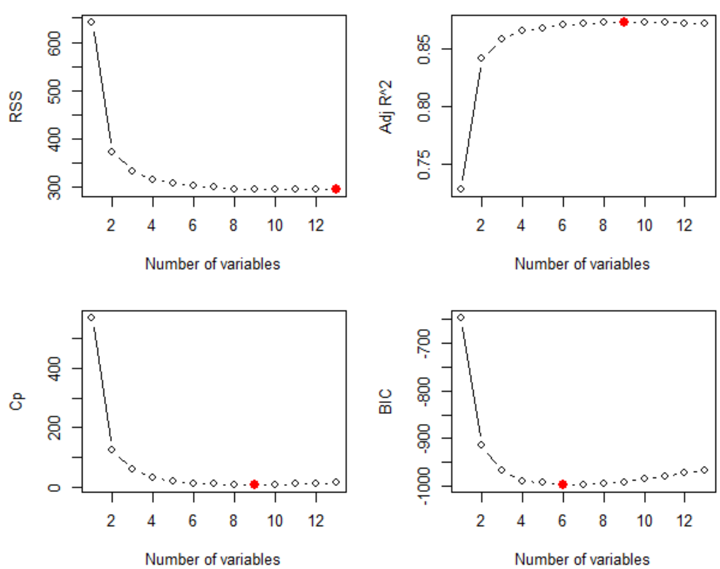

Figure 1: **Modelo *Regsubset, method Exhautive.***

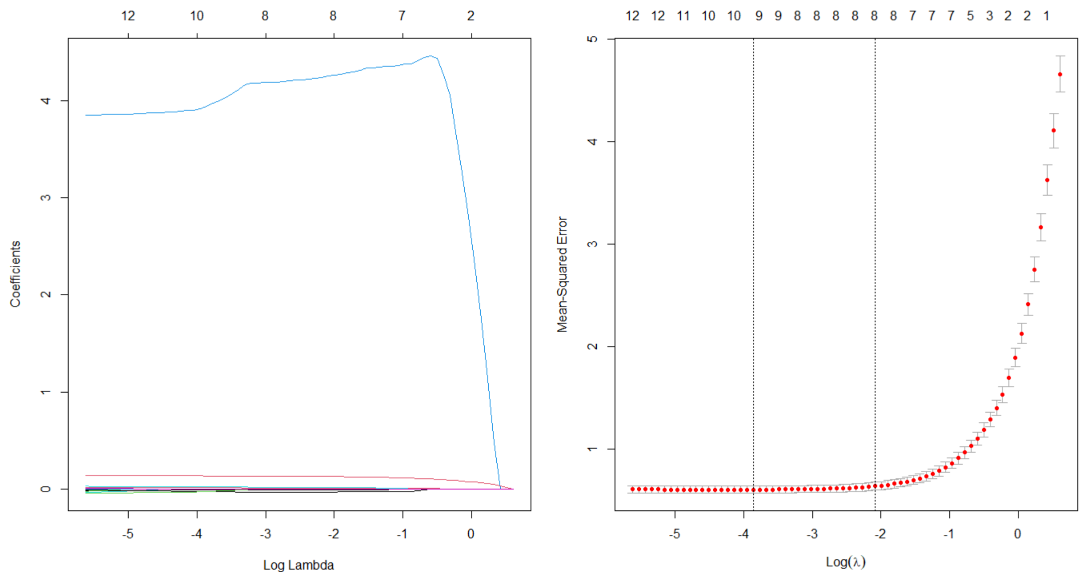

Figure 2: **Modelo *lasso.***

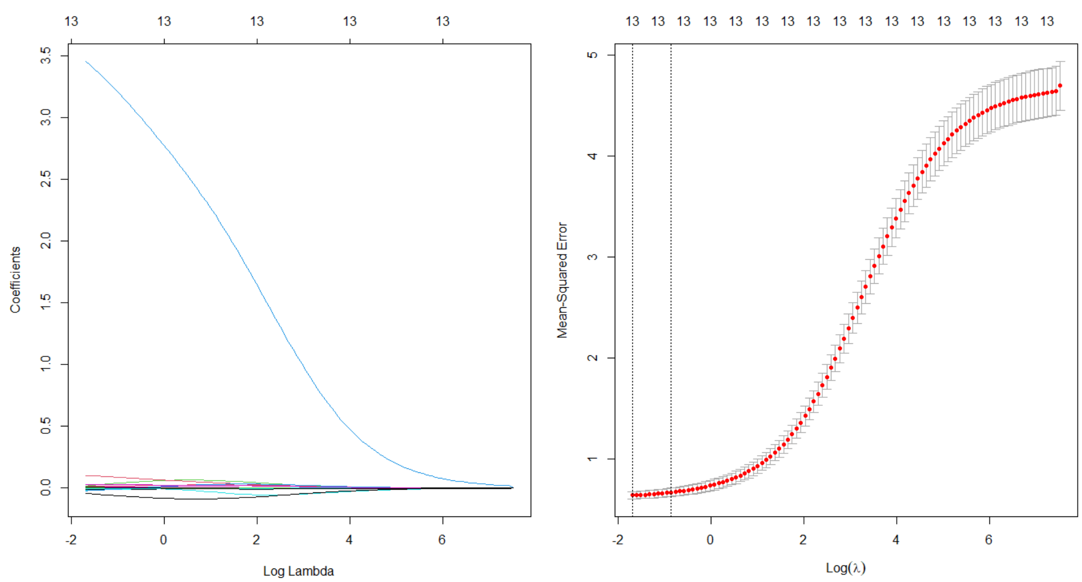

Figure 3: **Modelo *ridge.***

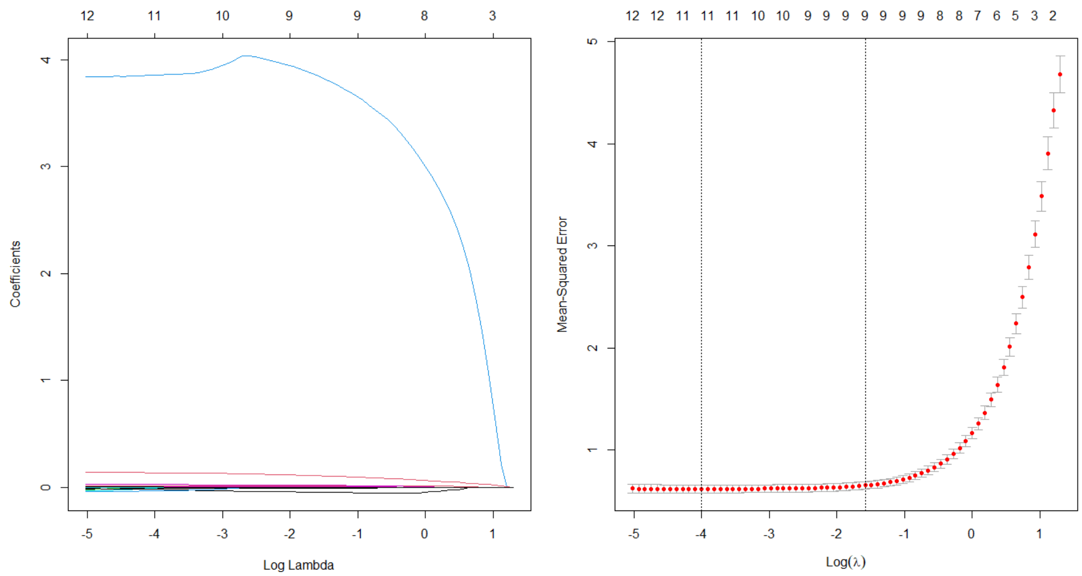

Figure 4: **Modelo *elastic net* con** `alpha = 0.5`

Ejercicio 2 Para los modelos que seleccionamos usando el criterio de λ_{min} calculamos el número de coeficientes que en valor absoluto son mayores 0.025:

ridge: 6		*lasso:* 29		*elastic net:* 42	
xLYSC_at	-0.033	xYOAB_at	-0.784	xYOAB_at	-0.563
xYEBC_at	-0.032	**xYEBC_at**	-0.531	**xYEBC_at**	-0.497
xYFIT_at	0.032	xLYSC_at	-0.270	xLYSC_at	-0.396
xSPOVAA_at	0.028	xSPOVAA_at	0.247	xSPOVAA_at	0.283
xYBFI_at	0.027	xYQJU_at	0.213	xYQJU_at	0.198
xHUTP_at	-0.027	xYXLD_at	0.197	xYBFI_at	0.173

Table 8: **Coeficientes con mayor valor absoluto** para cada uno de los modelos y su variable asociada.

	ridge	*lasso*	*elastic net*
λ_{minECM}	5.93	0.042	0.069
Num. betas (λ_{minECM})	4089	38	57
λ_{1se}	30.22	0.080	0.161
Num. betas (λ_{1se})	4089	29	42
λ_{maxdf}	-	0.006	0.012
Num. betas (λ_{maxdf})	-	72	90

Table 9: **Información adicional de los modelos.** Valores de λ que cumplen tres criterios (`lambda.min, lambda.1se, dfmax`) y su correspondiente número de coeficientes que son distintos de cero.

Para los 3 modelos es posible notar que el valor de λ que minimiza el error cuadrático medio no restringe demasiado a las β siendo en el primero en el que tienen más libertad.

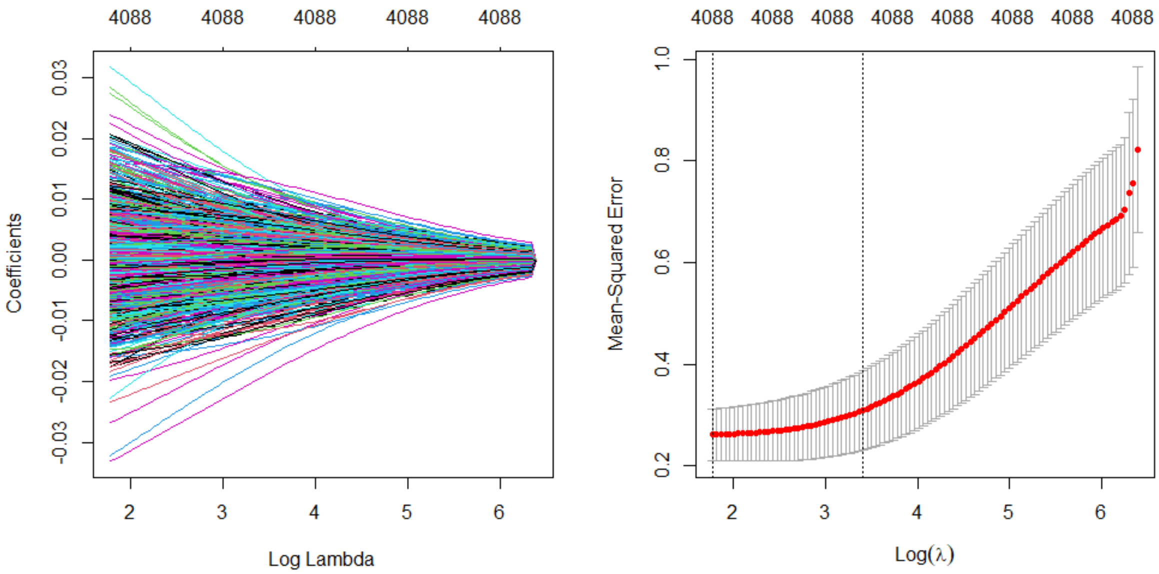

Figure 5: **Modelo *ridge*.**

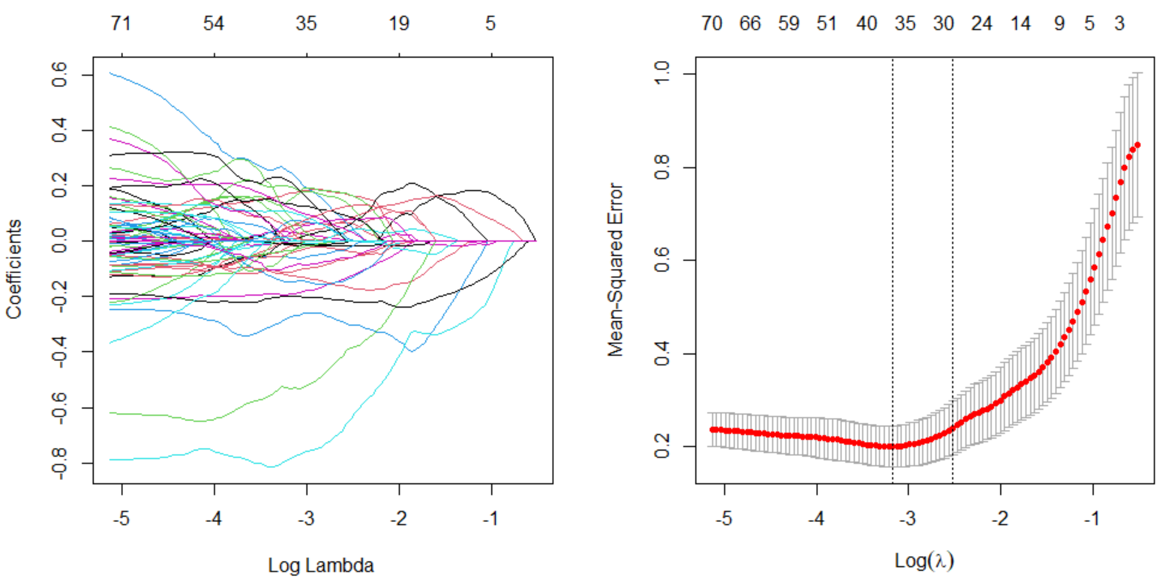

Figure 6: **Modelo *lasso*.**

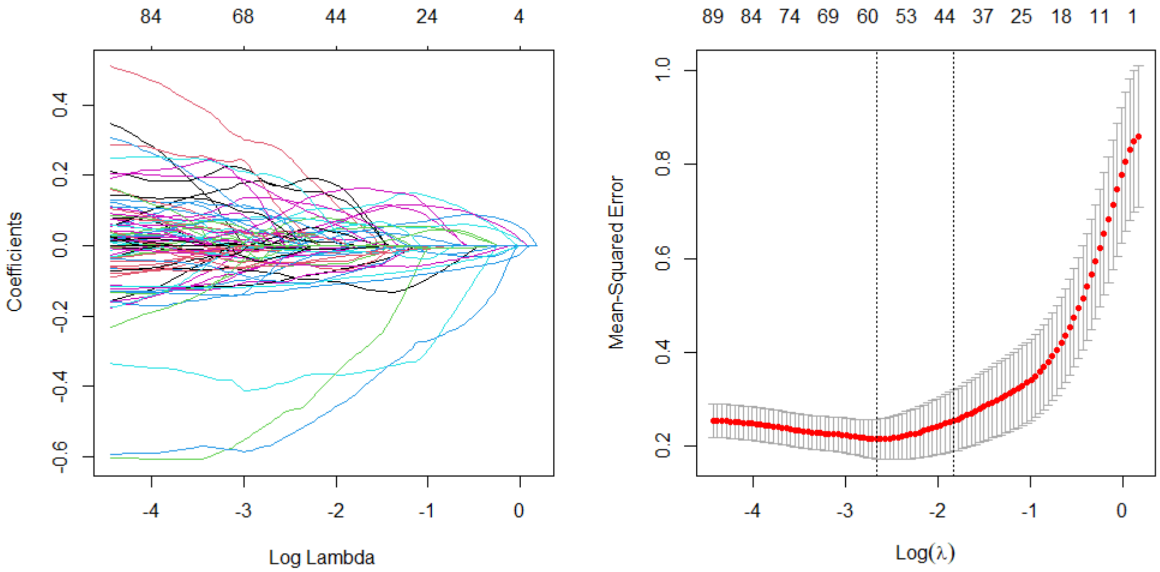

Figure 7: **Modelo *elastic net* con** `alpha = 0.5`