TRABAJO DE EVALUACIÓN (Regresión con regularización)

María Marín Cerdá

Carga de datos

```
library(readr)
datos = read.csv("dataworkMASTER.csv",sep=";",header=TRUE)
dim(datos)
```

```
[1] 4000   42
```

```
names(datos)
```

```
 [1] "Cod_Id" "varobj" "x01"    "x02"    "x03"    "x04"    "x05"    "x06"
 [9] "x07"    "x08"    "x09"    "x10"    "x11"    "x12"    "x13"    "x14"
[17] "x15"    "x16"    "x17"    "x18"    "x19"    "x20"    "x21"    "x22"
[25] "x23"    "x24"    "x25"    "x26"    "x27"    "x28"    "x29"    "x30"
[33] "x31"    "x32"    "x33"    "x34"    "x35"    "x36"    "x37"    "x38"
[41] "x39"    "x40"
```

```
str(datos)
```

```
'data.frame':   4000 obs. of  42 variables:
 $ Cod_Id: int  1 2 3 4 5 6 7 8 9 10 ...
 $ varobj: int  419 407 438 430 408 419 406 424 392 402 ...
 $ x01   : int  9 10 13 9 6 13 8 7 11 11 ...
 $ x02   : int  19 20 20 16 22 20 21 22 17 16 ...
 $ x03   : int  20 18 20 22 19 20 21 21 18 20 ...
 $ x04   : int  37 38 46 37 31 41 32 33 39 41 ...
 $ x05   : int  58 57 58 48 62 55 60 61 53 50 ...
```

```
 $ x06   : int  56 49 53 63 53 57 57 59 55 55 ...
 $ x07   : int  86 90 105 86 73 97 78 81 90 95 ...
 $ x08   : int  21 21 23 22 18 20 19 23 17 18 ...
 $ x09   : int  22 19 21 24 21 20 19 24 15 20 ...
 $ x10   : int  55 58 65 65 56 54 57 64 53 54 ...
 $ x11   : int  59 58 62 64 55 61 56 62 55 52 ...
 $ x12   : int  37 32 32 44 38 34 34 35 27 33 ...
 $ x13   : int  67 69 77 77 68 69 75 83 63 69 ...
 $ x14   : int  73 72 79 79 70 74 73 79 74 66 ...
 $ x15   : int  21 19 17 16 19 19 19 20 22 19 ...
 $ x16   : int  24 17 21 20 20 21 20 18 17 19 ...
 $ x17   : int  18 19 21 19 19 20 18 17 19 18 ...
 $ x18   : int  44 43 40 38 42 42 41 40 43 41 ...
 $ x19   : int  45 41 41 42 44 39 44 37 40 39 ...
 $ x20   : int  46 42 45 44 42 44 42 38 43 44 ...
 $ x21   : int  58 58 59 55 59 57 57 58 57 56 ...
 $ x22   : int  18 17 19 18 23 19 18 20 20 18 ...
 $ x23   : int  19 19 18 19 20 20 19 17 20 19 ...
 $ x24   : int  25 20 20 20 19 23 17 19 19 21 ...
 $ x25   : int  32 33 35 38 41 38 39 36 40 34 ...
 $ x26   : int  34 37 38 33 37 39 40 32 37 43 ...
 $ x27   : int  44 35 39 35 34 41 29 33 31 38 ...
 $ x28   : int  48 45 49 52 61 56 55 53 58 50 ...
 $ x29   : int  51 51 54 54 54 55 57 50 50 56 ...
 $ x30   : int  21 17 17 19 19 18 19 18 20 21 ...
 $ x31   : int  17 21 21 24 19 22 19 21 22 17 ...
 $ x32   : int  16 18 19 16 19 20 21 18 20 20 ...
 $ x33   : int  20 19 25 21 21 16 17 20 19 19 ...
 $ x34   : int  36 39 39 42 36 36 37 36 40 36 ...
 $ x35   : int  31 36 34 35 37 39 37 36 40 35 ...
 $ x36   : int  33 36 44 36 38 31 30 41 30 37 ...
 $ x37   : int  54 54 57 58 51 53 51 52 55 51 ...
 $ x38   : int  46 52 49 49 49 53 53 51 58 53 ...
 $ x39   : int  42 50 59 47 58 46 41 56 45 49 ...
 $ x40   : int  67 70 72 73 67 66 65 67 69 65 ...
```

```
datos=datos[,-1]  # Eliminación de la variable código Cod_Id
```

Preparación del conjunto de datos

```
a=3
b=0
c=5
d=9
elim=c(10+a, 10+b, 20+c, 30+d)
datos=datos[,-elim]
```

Trabajaremos con 4000 observaciones y 37 variables.

Ejercicio 1

> **Note**
>
> Obtener el modelo de regresión múltiple de la variable `varobj` frente al resto de variables. Analizar el problema de multicolinealidad.

Construimos el modelo de regresión lineal múltiple.

```
regre<-lm(varobj~.,data=datos)
summary(regre)
```

```

Call:
lm(formula = varobj ~ ., data = datos)

Residuals:
     Min       1Q   Median       3Q      Max 
-12.9134  -2.3894  -0.0808   2.3517  12.3507 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.400e+02  3.213e+00  43.583  < 2e-16 ***
x01          1.073e+00  6.298e-02  17.039  < 2e-16 ***
x02         -1.685e-02  6.235e-02  -0.270 0.786951    
x03          1.078e+00  6.111e-02  17.645  < 2e-16 ***
x04          1.026e+00  6.291e-02  16.303  < 2e-16 ***
x05          9.968e-01  2.796e-02  35.647  < 2e-16 ***
x06          9.655e-01  2.745e-02  35.170  < 2e-16 ***
```

```
x07         -7.521e-02  2.777e-02  -2.708 0.006788 **
x08          1.356e+00  8.339e-02  16.264  < 2e-16 ***
x10          7.798e-01  3.866e-02  20.168  < 2e-16 ***
x11          1.110e-01  3.951e-02   2.809 0.005001 **
x13          3.371e-02  2.761e-02   1.221 0.222139
x14         -1.026e-01  2.758e-02  -3.718 0.000203 ***
x15         -7.577e-02  3.852e-02  -1.967 0.049272 *
x16          9.716e-01  3.935e-02  24.691  < 2e-16 ***
x17          2.060e-02  4.030e-02   0.511 0.609247
x18          1.989e-02  3.909e-02   0.509 0.610877
x19         -3.799e-02  2.768e-02  -1.372 0.170017
x20         -1.092e-03  2.768e-02  -0.039 0.968542
x21         -9.846e-02  2.756e-02  -3.573 0.000357 ***
x22         -3.001e-02  3.857e-02  -0.778 0.436562
x23          7.423e-02  3.877e-02   1.914 0.055643 .
x25         -2.296e-03  3.874e-02  -0.059 0.952753
x26          2.920e-02  3.940e-02   0.741 0.458656
x27          1.185e-01  2.002e-02   5.920 3.49e-09 ***
x28         -4.612e-04  2.750e-02  -0.017 0.986619
x29         -5.509e-02  2.758e-02  -1.997 0.045865 *
x30          1.061e-01  2.751e-02   3.858 0.000116 ***
x31         -4.004e-02  3.904e-02  -1.026 0.305133
x32          1.356e-01  3.945e-02   3.437 0.000593 ***
x33          1.050e-01  3.897e-02   2.694 0.007087 **
x34         -4.926e-02  3.903e-02  -1.262 0.206979
x35         -1.201e-01  2.842e-02  -4.228 2.41e-05 ***
x36          6.918e-02  3.931e-02   1.760 0.078533 .
x37         -4.741e-02  3.839e-02  -1.235 0.216921
x39          1.317e-01  2.789e-02   4.723 2.41e-06 ***
x40          9.761e-02  2.725e-02   3.582 0.000345 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.521 on 3963 degrees of freedom
Multiple R-squared:  0.9138,    Adjusted R-squared:  0.913
F-statistic:  1167 on 36 and 3963 DF,  p-value: < 2.2e-16
```

Observamos que existen variables no significativas para el modelo y que el valor de R^2 es bastante alto. Estudiaremos si es debido a que el modelo es bastante bueno o si existe un problema de multicolinealidad.

```r
str(datos)
```

```
'data.frame':	4000 obs. of  37 variables:
 $ varobj: int  419 407 438 430 408 419 406 424 392 402 ...
 $ x01   : int  9 10 13 9 6 13 8 7 11 11 ...
 $ x02   : int  19 20 20 16 22 20 21 22 17 16 ...
 $ x03   : int  20 18 20 22 19 20 21 21 18 20 ...
 $ x04   : int  37 38 46 37 31 41 32 33 39 41 ...
 $ x05   : int  58 57 58 48 62 55 60 61 53 50 ...
 $ x06   : int  56 49 53 63 53 57 57 59 55 55 ...
 $ x07   : int  86 90 105 86 73 97 78 81 90 95 ...
 $ x08   : int  21 21 23 22 18 20 19 23 17 18 ...
 $ x10   : int  55 58 65 65 56 54 57 64 53 54 ...
 $ x11   : int  59 58 62 64 55 61 56 62 55 52 ...
 $ x13   : int  67 69 77 77 68 69 75 83 63 69 ...
 $ x14   : int  73 72 79 79 70 74 73 79 74 66 ...
 $ x15   : int  21 19 17 16 19 19 19 20 22 19 ...
 $ x16   : int  24 17 21 20 20 21 20 18 17 19 ...
 $ x17   : int  18 19 21 19 19 20 18 17 19 18 ...
 $ x18   : int  44 43 40 38 42 42 41 40 43 41 ...
 $ x19   : int  45 41 41 42 44 39 44 37 40 39 ...
 $ x20   : int  46 42 45 44 42 44 42 38 43 44 ...
 $ x21   : int  58 58 59 55 59 57 57 58 57 56 ...
 $ x22   : int  18 17 19 18 23 19 18 20 20 18 ...
 $ x23   : int  19 19 18 19 20 20 19 17 20 19 ...
 $ x25   : int  32 33 35 38 41 38 39 36 40 34 ...
 $ x26   : int  34 37 38 33 37 39 40 32 37 43 ...
 $ x27   : int  44 35 39 35 34 41 29 33 31 38 ...
 $ x28   : int  48 45 49 52 61 56 55 53 58 50 ...
 $ x29   : int  51 51 54 54 54 55 57 50 50 56 ...
 $ x30   : int  21 17 17 19 19 18 19 18 20 21 ...
 $ x31   : int  17 21 21 24 19 22 19 21 22 17 ...
 $ x32   : int  16 18 19 16 19 20 21 18 20 20 ...
 $ x33   : int  20 19 25 21 21 16 17 20 19 19 ...
 $ x34   : int  36 39 39 42 36 36 37 36 40 36 ...
 $ x35   : int  31 36 34 35 37 39 37 36 40 35 ...
 $ x36   : int  33 36 44 36 38 31 30 41 30 37 ...
 $ x37   : int  54 54 57 58 51 53 51 52 55 51 ...
 $ x39   : int  42 50 59 47 58 46 41 56 45 49 ...
 $ x40   : int  67 70 72 73 67 66 65 67 69 65 ...
```

Como todas las variables son cuantitativas, podemos calcular la matriz de correlación.

```
mx<-datos[,-1]

(mat_cor<-cor(mx))
```

```
             x01           x02           x03          x04          x05
x01  1.000000000 -4.619657e-03  0.0312951258  0.900631212 -0.017682095
x02 -0.004619657  1.000000e+00  0.0007784952 -0.051634371  0.890352587
x03  0.031295126  7.784952e-04  1.0000000000  0.023665042 -0.044046252
x04  0.900631212 -5.163437e-02  0.0236650421  1.000000000 -0.059982341
x05 -0.017682095  8.903526e-01 -0.0440462519 -0.059982341  1.000000000
x06 -0.047334482  3.181154e-03  0.8900246394 -0.066465474 -0.032028109
x07  0.875696829 -8.966602e-02  0.0274753616  0.975474191 -0.104016530
x08 -0.005958752  9.478935e-05 -0.0139443783 -0.011296242  0.011335375
x10 -0.013911426  1.210211e-02 -0.0313463477 -0.020481144  0.023649052
x11 -0.008979966 -5.177154e-03 -0.0127286775 -0.009783947  0.005542555
x13 -0.010598345  1.761031e-02 -0.0289551141 -0.018309951  0.025312423
x14 -0.013033578 -7.226456e-03 -0.0101490923 -0.009020713 -0.001099545
x15 -0.021680150  8.073577e-03 -0.0028878710 -0.013406602  0.012827950
x16 -0.017522115 -1.344206e-02 -0.0190660671 -0.017134710 -0.014767217
x17  0.038762445  1.280565e-02  0.0083759627  0.025465131  0.011311559
x18 -0.010690256  2.654630e-02  0.0084587379 -0.006985097  0.024021385
x19 -0.029795412 -2.555600e-02 -0.0048654742 -0.026534255 -0.030259157
x20  0.038556423 -5.895913e-03  0.0166097772  0.034021658  0.002544535
x21  0.005070909  3.384388e-02  0.0078062132  0.002011747  0.027285482
x22 -0.021892479 -7.681681e-03 -0.0097073106 -0.017238706 -0.007159560
x23  0.009427539  1.303405e-02 -0.0146085667  0.014816525  0.005412979
x25 -0.020627632  1.854368e-02  0.0053665219 -0.017729359  0.012374307
x26 -0.008536652 -1.125881e-02 -0.0414858938 -0.001233093 -0.018815757
x27 -0.012303811 -2.171440e-02 -0.0077419104  0.001189883 -0.018197203
x28 -0.001835559 -7.218417e-03  0.0197604238 -0.001247325 -0.009227945
x29 -0.011468353  5.204864e-03 -0.0291433756 -0.007171973  0.003526424
x30 -0.035902389  2.963704e-03 -0.0132480384 -0.037806922 -0.002058595
x31  0.006486975  1.466364e-02 -0.0006877174  0.013741412  0.022340311
x32 -0.006320284 -1.618230e-02 -0.0151147222 -0.005037558 -0.015177826
x33 -0.025728860  8.915054e-03  0.0231532117 -0.018459550  0.014018565
x34 -0.002709053 -1.659841e-03  0.0097849946  0.008196896  0.002533485
x35  0.008596538 -6.934421e-03 -0.0087025164  0.002431423 -0.008742193
x36 -0.017216008  5.387077e-03  0.0286573602 -0.014565081  0.004571477
x37 -0.013691737 -5.621005e-03  0.0179490079 -0.010015408 -0.007379240
x39 -0.012185233  1.723853e-02  0.0149934763 -0.013379235  0.009050623
x40 -0.018994216  7.922424e-04  0.0113530637 -0.012397451 -0.008211633
              x06           x07           x08           x10          x11
```

x01	-0.0473344819	0.8756968293	-5.958752e-03	-0.0139114263	-0.008979966
x02	0.0031811540	-0.0896660162	9.478935e-05	0.0121021057	-0.005177154
x03	0.8900246394	0.0274753616	-1.394438e-02	-0.0313463477	-0.012728677
x04	-0.0664654740	0.9754741910	-1.129624e-02	-0.0204811437	-0.009783947
x05	-0.0320281086	-0.1040165297	1.133537e-02	0.0236490519	0.005542555
x06	1.0000000000	-0.0634292781	-2.311655e-02	-0.0329441524	-0.019307014
x07	-0.0634292781	1.0000000000	-1.243073e-02	-0.0233929328	-0.012207560
x08	-0.0231165483	-0.0124307307	1.000000e+00	0.8926235665	0.893535390
x10	-0.0329441524	-0.0233929328	8.926236e-01	1.0000000000	0.795778437
x11	-0.0193070143	-0.0122075596	8.935354e-01	0.7957784367	1.000000000
x13	-0.0327263528	-0.0216950670	7.875470e-01	0.9005890684	0.679284232
x14	-0.0151094825	-0.0100015922	8.156683e-01	0.7281692916	0.912820117
x15	-0.0080072770	-0.0180294305	1.204681e-03	0.0042774354	-0.006323996
x16	-0.0168595164	-0.0188376016	-3.776874e-03	-0.0073991701	-0.008366205
x17	0.0135901727	0.0262972912	-3.997643e-02	-0.0264413603	-0.043675360
x18	0.0036525569	-0.0083471484	1.651370e-02	0.0109449604	0.008773031
x19	0.0035036005	-0.0293143060	-1.213042e-02	-0.0201749504	-0.009521720
x20	0.0102377489	0.0362420231	-3.780088e-02	-0.0269949270	-0.048149840
x21	0.0009473664	0.0016389307	2.690420e-02	0.0241592542	0.015401359
x22	-0.0045571134	-0.0143631596	-1.248081e-02	-0.0119586891	-0.021379080
x23	-0.0112313013	0.0149291746	-4.087137e-03	-0.0003669357	0.003927104
x25	0.0020631940	-0.0181194229	-6.290866e-03	-0.0033906994	-0.011700069
x26	-0.0278611563	0.0007020763	-1.498504e-02	-0.0132829198	-0.014490812
x27	-0.0233242591	0.0072378116	1.853231e-02	0.0250154174	0.015529206
x28	0.0184182235	-0.0042714044	2.571407e-03	0.0045125629	0.002837105
x29	-0.0167740278	-0.0062021298	-9.419155e-03	-0.0066770425	-0.008022317
x30	-0.0064617523	-0.0346316549	-1.648271e-02	-0.0082020099	-0.006724378
x31	0.0093743383	0.0085061267	2.130486e-02	0.0142527024	0.026414331
x32	-0.0224749385	-0.0020381318	-4.499618e-02	-0.0339469363	-0.040018850
x33	0.0158098612	-0.0186794692	-1.456028e-02	-0.0144256164	-0.016008508
x34	0.0062334100	0.0054223543	3.275327e-02	0.0239059499	0.037516997
x35	-0.0097291154	0.0076974238	-4.381099e-02	-0.0346291730	-0.044588634
x36	0.0266113079	-0.0150270956	-1.559026e-02	-0.0232442752	-0.013410708
x37	0.0162711608	-0.0113399482	2.063617e-02	0.0175703632	0.024055736
x39	0.0101828757	-0.0143496860	-1.156845e-02	-0.0181600507	-0.013315594
x40	0.0072080050	-0.0138273145	1.152729e-02	0.0126685051	0.019694085

	x13	x14	x15	x16	x17
x01	-0.0105983449	-0.013033578	-0.021680150	-0.0175221150	0.038762445
x02	0.0176103112	-0.007226456	0.008073577	-0.0134420573	0.012805653
x03	-0.0289551141	-0.010149092	-0.002887871	-0.0190660671	0.008375963
x04	-0.0183099514	-0.009020713	-0.013406602	-0.0171347099	0.025465131
x05	0.0253124229	-0.001099545	0.012827950	-0.0147672168	0.011311559
x06	-0.0327263528	-0.015109483	-0.008007277	-0.0168595164	0.013590173

x07	-0.0216950670	-0.010001592	-0.018029431	-0.0188376016	0.026297291
x08	0.7875470073	0.815668309	0.001204681	-0.0037768743	-0.039976426
x10	0.9005890684	0.728169292	0.004277435	-0.0073991701	-0.026441360
x11	0.6792842324	0.912820117	-0.006323996	-0.0083662045	-0.043675360
x13	1.0000000000	0.623317989	0.004882035	-0.0027111559	-0.022010418
x14	0.6233179887	1.000000000	-0.003832194	0.0085948707	-0.041265377
x15	0.0048820354	-0.003832194	1.000000000	0.0156471749	0.014080947
x16	-0.0027111559	0.008594871	0.015647175	1.0000000000	-0.016496172
x17	-0.0220104182	-0.041265377	0.014080947	-0.0164961721	1.000000000
x18	0.0150699493	0.011119019	0.699033589	0.0824456579	0.008762621
x19	-0.0224704435	0.002106080	0.005987061	0.6962822453	0.091665935
x20	-0.0183984040	-0.046185721	0.081777347	0.0079752940	0.703998874
x21	0.0230694313	0.013884571	0.556932034	0.0170951127	-0.003357243
x22	-0.0106163378	-0.015632635	-0.005167069	0.0075702064	-0.021456676
x23	-0.0030176203	-0.005275169	0.012333362	-0.0021336427	-0.011535751
x25	0.0051317179	-0.006741597	0.015357314	-0.0019071887	-0.004146063
x26	-0.0075251645	-0.012453115	0.014201118	-0.0056173818	-0.011911509
x27	0.0202707952	0.022155769	0.012879166	0.0023793076	0.015116054
x28	0.0088914421	0.003406057	-0.004346584	0.0078340897	0.014470690
x29	-0.0025063034	-0.005163194	0.008791253	0.0001474349	-0.016158947
x30	-0.0006120272	-0.007092174	0.007627130	-0.0103004115	-0.013966984
x31	0.0183435841	0.020994650	0.006353563	0.0117487243	0.029159737
x32	-0.0365139778	-0.033819488	0.007660272	-0.0006626189	-0.003509417
x33	-0.0041280938	-0.022298604	-0.012401168	-0.0024589100	0.023715107
x34	0.0220542649	0.023180605	0.003510316	0.0054658141	0.044756013
x35	-0.0384848311	-0.040086157	-0.006389467	0.0042100678	-0.004708782
x36	-0.0169081004	-0.022987502	-0.013061092	-0.0166918908	0.001778503
x37	0.0110657383	0.016821354	0.001042982	-0.0063006143	0.055732455
x39	-0.0146336639	-0.016277138	-0.011791894	-0.0033359019	-0.008943653
x40	0.0089015810	0.012990803	0.010102663	-0.0147505917	0.046207455

	x18	x19	x20	x21	x22
x01	-0.0106902564	-0.0297954125	0.0385564235	0.0050709087	-0.0218924786
x02	0.0265462955	-0.0255560027	-0.0058959133	0.0338438830	-0.0076816815
x03	0.0084587379	-0.0048654742	0.0166097772	0.0078062132	-0.0097073106
x04	-0.0069850974	-0.0265342550	0.0340216579	0.0020117474	-0.0172387065
x05	0.0240213848	-0.0302591571	0.0025445349	0.0272854819	-0.0071595603
x06	0.0036525569	0.0035036005	0.0102377489	0.0009473664	-0.0045571134
x07	-0.0083471484	-0.0293143060	0.0362420231	0.0016389307	-0.0143631596
x08	0.0165136963	-0.0121304249	-0.0378008809	0.0269041988	-0.0124808138
x10	0.0109449604	-0.0201749504	-0.0269949270	0.0241592542	-0.0119586891
x11	0.0087730305	-0.0095217204	-0.0481498398	0.0154013593	-0.0213790804
x13	0.0150699493	-0.0224704435	-0.0183984040	0.0230694313	-0.0106163378
x14	0.0111190193	0.0021060802	-0.0461857214	0.0138845712	-0.0156326351

x15	0.6990335886	0.0059870610	0.0817773471	0.5569320344	-0.0051670686
x16	0.0824456579	0.6962822453	0.0079752940	0.0170951127	0.0075702064
x17	0.0087626210	0.0916659349	0.7039988739	-0.0033572426	-0.0214566757
x18	1.0000000000	0.0564634195	0.1042559927	0.8083180902	-0.0208420599
x19	0.0564634195	1.0000000000	0.0630154165	-0.0432228131	0.0178706411
x20	0.1042559927	0.0630154165	1.0000000000	0.0688107801	-0.0150591401
x21	0.8083180902	-0.0432228131	0.0688107801	1.0000000000	-0.0259063919
x22	-0.0208420599	0.0178706411	-0.0150591401	-0.0259063919	1.0000000000
x23	0.0091492743	0.0077860155	0.0001133584	-0.0033571005	0.0177000279
x25	-0.0049628133	0.0040793757	-0.0031931391	-0.0053344187	0.6847781596
x26	0.0029863882	-0.0019384917	-0.0004866685	-0.0085706423	-0.0058195774
x27	0.0060633110	-0.0045813768	0.0168803594	-0.0036914815	-0.0674319475
x28	-0.0122976178	0.0136914969	0.0095991440	-0.0192043548	0.5537273544
x29	0.0126785821	0.0008741468	0.0068883906	-0.0055258178	0.0098777144
x30	-0.0081545538	-0.0253060450	-0.0027502366	0.0030413349	-0.0150738612
x31	0.0080841164	0.0155761421	0.0261562775	0.0062673677	0.0126122808
x32	-0.0269494173	-0.0040769839	-0.0073097600	-0.0086455402	-0.0141197800
x33	-0.0037398382	0.0005409523	0.0173895410	0.0029217355	0.0195017392
x34	0.0004058703	0.0084221229	0.0456409942	0.0018960142	-0.0158599208
x35	-0.0333094777	-0.0007898042	-0.0036852384	-0.0118406733	-0.0166162987
x36	-0.0098109447	-0.0107584263	-0.0089931861	-0.0089161570	0.0118268081
x37	0.0065574901	-0.0035930539	0.0550727510	0.0052895614	-0.0017942988
x39	0.0046487155	-0.0036184618	-0.0091642677	-0.0009771870	0.0157571299
x40	0.0206877301	-0.0174621560	0.0467079777	0.0194412906	-0.0005029804

	x23	x25	x26	x27	x28
x01	0.0094275393	-0.020627632	-0.0085366521	-0.012303811	-0.001835559
x02	0.0130340531	0.018543685	-0.0112588112	-0.021714396	-0.007218417
x03	-0.0146085667	0.005366522	-0.0414858938	-0.007741910	0.019760424
x04	0.0148165249	-0.017729359	-0.0012330929	0.001189883	-0.001247325
x05	0.0054129785	0.012374307	-0.0188157572	-0.018197203	-0.009227945
x06	-0.0112313013	0.002063194	-0.0278611563	-0.023324259	0.018418224
x07	0.0149291746	-0.018119423	0.0007020763	0.007237812	-0.004271404
x08	-0.0040871366	-0.006290866	-0.0149850372	0.018532310	0.002571407
x10	-0.0003669357	-0.003390699	-0.0132829198	0.025015417	0.004512563
x11	0.0039271044	-0.011700069	-0.0144908124	0.015529206	0.002837105
x13	-0.0030176203	0.005131718	-0.0075251645	0.020270795	0.008891442
x14	-0.0052751689	-0.006741597	-0.0124531151	0.022155769	0.003406057
x15	0.0123333617	0.015357314	0.0142011179	0.012879166	-0.004346584
x16	-0.0021336427	-0.001907189	-0.0056173818	0.002379308	0.007834090
x17	-0.0115357505	-0.004146063	-0.0119115091	0.015116054	0.014470690
x18	0.0091492743	-0.004962813	0.0029863882	0.006063311	-0.012297618
x19	0.0077860155	0.004079376	-0.0019384917	-0.004581377	0.013691497
x20	0.0001133584	-0.003193139	-0.0004866685	0.016880359	0.009599144

x21	-0.0033571005	-0.005334419	-0.0085706423	-0.003691482	-0.019204355
x22	0.0177000279	0.684778160	-0.0058195774	-0.067431947	0.553727354
x23	1.0000000000	-0.055662905	0.7025389009	0.005515396	-0.110459428
x25	-0.0556629048	1.000000000	-0.0499756899	-0.094682277	0.810489993
x26	0.7025389009	-0.049975690	1.0000000000	-0.058666568	-0.130269780
x27	0.0055153958	-0.094682277	-0.0586665684	1.000000000	-0.061882953
x28	-0.1104594284	0.810489993	-0.1302697801	-0.061882953	1.000000000
x29	0.5711742055	-0.028283776	0.8178264918	-0.137899026	-0.092973199
x30	-0.0112158495	0.001779216	0.0045771089	0.003151472	0.021629845
x31	-0.0210112491	0.025231588	-0.0229761689	0.008078299	-0.001833290
x32	-0.0086646775	0.002021937	0.0032190540	0.017737960	0.018404469
x33	0.0112340468	0.001552755	0.0168954423	0.019472481	0.007662392
x34	-0.0104632143	-0.004638422	-0.0190723606	0.007779651	-0.018687024
x35	-0.0225274010	-0.012524039	-0.0057863332	0.006385163	0.007459952
x36	0.0215907994	-0.002221458	0.0246000011	0.017437300	0.010622327
x37	0.0043178251	0.017340378	-0.0218108243	0.007163400	0.005335063
x39	0.0183109450	0.003864152	0.0342380016	0.018171156	0.014004357
x40	0.0092858776	0.007009480	-0.0237249468	0.006212083	-0.012855761

	x29	x30	x31	x32	x33
x01	-0.0114683526	-0.0359023885	0.0064869749	-0.0063202841	-0.0257288596
x02	0.0052048640	0.0029637036	0.0146636352	-0.0161822955	0.0089150540
x03	-0.0291433756	-0.0132480384	-0.0006877174	-0.0151147222	0.0231532117
x04	-0.0071719731	-0.0378069217	0.0137414121	-0.0050375584	-0.0184595495
x05	0.0035264238	-0.0020585945	0.0223403112	-0.0151778258	0.0140185649
x06	-0.0167740278	-0.0064617523	0.0093743383	-0.0224749385	0.0158098612
x07	-0.0062021298	-0.0346316549	0.0085061267	-0.0020381318	-0.0186794692
x08	-0.0094191553	-0.0164827067	0.0213048648	-0.0449961849	-0.0145602834
x10	-0.0066770425	-0.0082020099	0.0142527024	-0.0339469363	-0.0144256164
x11	-0.0080223165	-0.0067243784	0.0264143311	-0.0400188497	-0.0160085077
x13	-0.0025063034	-0.0006120272	0.0183435841	-0.0365139778	-0.0041280938
x14	-0.0051631941	-0.0070921740	0.0209946499	-0.0338194884	-0.0222986039
x15	0.0087912526	0.0076271298	0.0063535626	0.0076602718	-0.0124011676
x16	0.0001474349	-0.0103004115	0.0117487243	-0.0006626189	-0.0024589100
x17	-0.0161589467	-0.0139669842	0.0291597370	-0.0035094172	0.0237151071
x18	0.0126785821	-0.0081545538	0.0080841164	-0.0269494173	-0.0037398382
x19	0.0008741468	-0.0253060450	0.0155761421	-0.0040769839	0.0005409523
x20	0.0068883906	-0.0027502366	0.0261562775	-0.0073097600	0.0173895410
x21	-0.0055258178	0.0030413349	0.0062673677	-0.0086455402	0.0029217355
x22	0.0098777144	-0.0150738612	0.0126122808	-0.0141197800	0.0195017392
x23	0.5711742055	-0.0112158495	-0.0210112491	-0.0086646775	0.0112340468
x25	-0.0282837761	0.0017792163	0.0252315880	0.0020219374	0.0015527548
x26	0.8178264918	0.0045771089	-0.0229761689	0.0032190540	0.0168954423
x27	-0.1378990256	0.0031514717	0.0080782987	0.0177379603	0.0194724811

x28	-0.0929731986	0.0216298452	-0.0018332903	0.0184044688	0.0076623917
x29	1.0000000000	0.0051649656	-0.0157966472	-0.0148191772	0.0269217614
x30	0.0051649656	1.0000000000	-0.0011299340	0.0161771093	-0.0100882207
x31	-0.0157966472	-0.0011299340	1.0000000000	-0.0024708422	0.0234523860
x32	-0.0148191772	0.0161771093	-0.0024708422	1.0000000000	0.0082574708
x33	0.0269217614	-0.0100882207	0.0234523860	0.0082574708	1.0000000000
x34	-0.0085439245	-0.0089235678	0.7032895812	-0.0861467704	0.0090365900
x35	-0.0149338265	0.0061380064	-0.0138544388	0.7141610011	-0.0731826451
x36	0.0255916679	-0.0102364215	-0.0501166194	0.0215979512	0.7030007276
x37	-0.0139754239	-0.0044473481	0.5695927447	-0.1191188798	0.0183404876
x39	0.0372502868	-0.0113675867	-0.0933502505	0.0250758556	0.5719863040
x40	-0.0164144823	0.0002891290	0.4850892891	-0.1611604156	0.0315571273

	x34	x35	x36	x37	x39
x01	-0.0027090531	0.0085965376	-0.017216008	-0.013691737	-0.012185233
x02	-0.0016598415	-0.0069344213	0.005387077	-0.005621005	0.017238530
x03	0.0097849946	-0.0087025164	0.028657360	0.017949008	0.014993476
x04	0.0081968958	0.0024314226	-0.014565081	-0.010015408	-0.013379235
x05	0.0025334845	-0.0087421932	0.004571477	-0.007379240	0.009050623
x06	0.0062334100	-0.0097291154	0.026611308	0.016271161	0.010182876
x07	0.0054223543	0.0076974238	-0.015027096	-0.011339948	-0.014349686
x08	0.0327532710	-0.0438109897	-0.015590256	0.020636172	-0.011568451
x10	0.0239059499	-0.0346291730	-0.023244275	0.017570363	-0.018160051
x11	0.0375169970	-0.0445886344	-0.013410708	0.024055736	-0.013315594
x13	0.0220542649	-0.0384848311	-0.016908100	0.011065738	-0.014633664
x14	0.0231806051	-0.0400861571	-0.022987502	0.016821354	-0.016277138
x15	0.0035103163	-0.0063894673	-0.013061092	0.001042982	-0.011791894
x16	0.0054658141	0.0042100678	-0.016691891	-0.006300614	-0.003335902
x17	0.0447560128	-0.0047087822	0.001778503	0.055732455	-0.008943653
x18	0.0004058703	-0.0333094777	-0.009810945	0.006557490	0.004648715
x19	0.0084221229	-0.0007898042	-0.010758426	-0.003593054	-0.003618462
x20	0.0456409942	-0.0036852384	-0.008993186	0.055072751	-0.009164268
x21	0.0018960142	-0.0118406733	-0.008916157	0.005289561	-0.000977187
x22	-0.0158599208	-0.0166162987	0.011826808	-0.001794299	0.015757130
x23	-0.0104632143	-0.0225274010	0.021590799	0.004317825	0.018310945
x25	-0.0046384218	-0.0125240388	-0.002221458	0.017340378	0.003864152
x26	-0.0190723606	-0.0057863332	0.024600001	-0.021810824	0.034238002
x27	0.0077796512	0.0063851626	0.017437300	0.007163400	0.018171156
x28	-0.0186870244	0.0074599520	0.010622327	0.005335063	0.014004357
x29	-0.0085439245	-0.0149338265	0.025591668	-0.013975424	0.037250287
x30	-0.0089235678	0.0061380064	-0.010236421	-0.004447348	-0.011367587
x31	0.7032895812	-0.0138544388	-0.050116619	0.569592745	-0.093350250
x32	-0.0861467704	0.7141610011	0.021597951	-0.119118880	0.025075856
x33	0.0090365900	-0.0731826451	0.703000728	0.018340488	0.571986304

```
x34  1.0000000000 -0.0571566997 -0.100207332  0.810157506 -0.157557378
x35 -0.0571566997  1.0000000000 -0.042957966 -0.121832838 -0.013606711
x36 -0.1002073321 -0.0429579662  1.000000000 -0.076596433  0.823383879
x37  0.8101575059 -0.1218328383 -0.076596433  1.000000000 -0.160475338
x39 -0.1575573776 -0.0136067108  0.823383879 -0.160475338  1.000000000
x40  0.7016890578 -0.1749677839 -0.056190460  0.864583221 -0.130858533
              x40
x01 -0.0189942156
x02  0.0007922424
x03  0.0113530637
x04 -0.0123974509
x05 -0.0082116332
x06  0.0072080050
x07 -0.0138273145
x08  0.0115272855
x10  0.0126685051
x11  0.0196940850
x13  0.0089015810
x14  0.0129908028
x15  0.0101026631
x16 -0.0147505917
x17  0.0462074550
x18  0.0206877301
x19 -0.0174621560
x20  0.0467079777
x21  0.0194412906
x22 -0.0005029804
x23  0.0092858776
x25  0.0070094803
x26 -0.0237249468
x27  0.0062120829
x28 -0.0128557611
x29 -0.0164144823
x30  0.0002891290
x31  0.4850892891
x32 -0.1611604156
x33  0.0315571273
x34  0.7016890578
x35 -0.1749677839
x36 -0.0561904596
x37  0.8645832206
x39 -0.1308585330
x40  1.0000000000
```

```r
det(mat_cor)
```

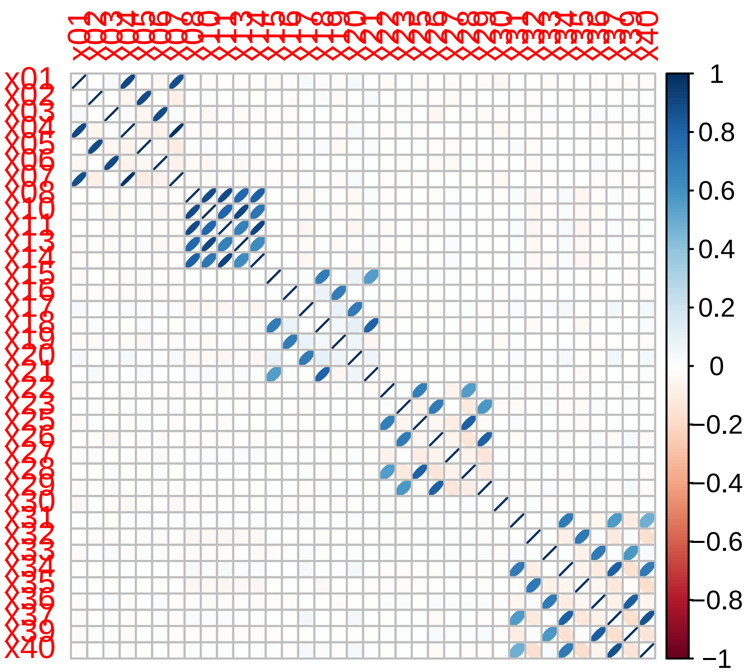
```
[1] 1.471774e-12
```

Observamos que existen varias variables altamente correladas y que el determinante de la matriz es muy cercano a 0. Veamos un gráfico para tener una idea más visual.

```r
library(corrplot)
```

```
corrplot 0.95 loaded
```

```r
corrplot(mat_cor,method="ellipse")
```



Efectivamente, como vimos en la matriz, existen variables altamente correladas como pueden ser las variables X_{01} y X_{04}; y otras, que en cambio, es prácticamente 0.

Otra medida para detectar multicolinealidad, es el coeficiente VIF (variance inflaction factor) de una variable predictora X_i. El VIF para una variable X_i se define como

$$VIF(X_i) = \frac{1}{1 - R_i^2}$$

siendo $R_i^2 = R^2[X_i; X_1, X_2, \dots, X_{i-1}, X_{i+1}, \dots, X_p]$, es decir, el coeficiente de determinación del modelo lineal de la variable X_i frente a las demás. Un valor R_i^2 alto, significará que la variable X_i es explicada por las demás. En conclusión, si VIF es alto (superior a 10) se dice que hay multicolinealidad muy alta.

```r
#install.packages("usdm")
library(usdm)
```

```
Warning: package 'usdm' was built under R version 4.4.2
```

```
Cargando paquete requerido: terra
```

```
terra 1.7.83
```

```r
vif(mx)
```

```
   Variables       VIF
1        x01  5.391316
2        x02  4.912256
3        x03  5.044263
4        x04 26.188206
5        x05  4.966944
6        x06  5.044783
7        x07 21.696898
8        x08  9.013989
9        x10 10.085224
10       x11 10.123466
11       x13  5.429661
12       x14  6.042500
13       x15  1.979925
14       x16  1.986168
15       x17  2.045683
16       x18  4.057524
17       x19  2.028658
18       x20  2.036944
19       x21  2.975309
20       x22  1.909408
21       x23  2.011623
22       x25  3.887107
23       x26  4.094784
24       x27  1.049806
```

```
25        x28  3.025873
26        x29  3.114539
27        x30  1.007511
28        x31  2.011115
29        x32  2.092356
30        x33  2.024825
31        x34  3.979001
32        x35  2.112653
33        x36  4.182798
34        x37  5.930525
35        x39  3.221988
36        x40  4.074512
```

Las variables que tienen un VIF superior a 10 son las variables X_{04}, X_{07}, X_{10} y X_{11} cuyos valores son 26.188206, 21.696898, 10.085224 y 10.123466 respectivamente.

Como hay multicolinealidad, es razonable aplicar métodos de regularización.

Ejercicio 2

> **Note**
>
> Aplicar la técnica de regularización elasticnet a través de la librería `glmnet`, para los valores de α en la colección $\{0, 0.1, 0.2, 0.3...0.9, 1.0\}$ seleccionando el mejor par (α,λ) por validación cruzada.

Procedamos a aplicar la técnica elasticnet, una técnica de mixtura entre la regularización Lasso y Ridge. Para ello utilizaremos la función `cv.glmnet` que realiza validación cruzada con k pliegues, proporciona un gráfico y un valor óptimo para λ dado un valor de α.

```r
set.seed(123)
#install.packages("glmnet")
library(glmnet)
```

```
Warning: package 'glmnet' was built under R version 4.4.2
```

```
Cargando paquete requerido: Matrix
```

```
Loaded glmnet 4.1-8
```

```r
mx<-as.matrix(mx)
my<-as.matrix(datos[,1])

mod_0<-cv.glmnet(mx,my,keep=TRUE,alpha=0)
```

Para $\alpha = 0$ la regularización que se aplica es la de Ridge. Por eso no se seleccionan variables y para todos los valores de λ los coeficientes de todas las variables son distinta de 0.

```r
mod_0$nzero
```

```
 s0  s1  s2  s3  s4  s5  s6  s7  s8  s9 s10 s11 s12 s13 s14 s15 s16 s17 s18 s19
 36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36
s20 s21 s22 s23 s24 s25 s26 s27 s28 s29 s30 s31 s32 s33 s34 s35 s36 s37 s38 s39
 36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36
s40 s41 s42 s43 s44 s45 s46 s47 s48 s49 s50 s51 s52 s53 s54 s55 s56 s57 s58 s59
 36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36
s60 s61 s62 s63 s64 s65 s66 s67 s68 s69 s70 s71 s72 s73 s74 s75 s76 s77 s78 s79
 36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36
s80 s81 s82 s83 s84 s85 s86 s87 s88 s89 s90 s91 s92 s93 s94 s95 s96 s97 s98 s99
 36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36  36
```

Se seleccionan las 36 variables independientes.

```r
mod_0
```

```
Call:  cv.glmnet(x = mx, y = my, keep = TRUE, alpha = 0)

Measure: Mean-Squared Error

    Lambda Index Measure     SE Nonzero
min 0.5911   100   12.98 0.1244      36
1se 0.7119    98   13.10 0.1271      36
```

Observamos que obtenemos $\lambda_{min} = 0.5911$ que minimiza el error con $MSE = 12.98$. El otro valor que aparece, $\lambda_{1se} = 0.7119$, es el λ más grande que está dentro de $\frac{+}{-}$ una desviación estándar del error mínimo. Es decir, que desde el punto de vista estadístico, el error es el mismo. Por lo tanto, seleccionaremos el valor de λ_{1se} pues al ser mayor reduce aún más la influencia de variables correladas y con el "mismo" error.

Vamos a crear un bucle que calcule el mejor λ para los distintos valores de α y tomaremos de nuevo λ_{1se}. Como α será mayor que 0, entrará en juego la regularización de tipo Lasso, y se procederá a seleccionar variables. El valor de λ_{1se} aportaría la misma información seleccionando menos variables por lo que reduciría la complejidad del modelo.

```
lista_alpha=(0)
lista_lambda=(mod_0$lambda.1se)
lista_error=(mod_0$cvm[mod_0$lambda == mod_0$lambda.1se])
for (a in seq(0.1,1,0.1)) {
  mod<-cv.glmnet(mx,my,keep=TRUE,alpha=a)
  lista_alpha<-append(lista_alpha,a)
  lista_lambda<-append(lista_lambda,mod$lambda.1se)
  lista_error<-append(lista_error,mod$cvm[mod$lambda == mod$lambda.1se])

}
info<-cbind(alpha=lista_alpha,lambda=lista_lambda,error=lista_error)
```

El valor de α que minimiza el error es 0.7. Procedamos a construir el modelo con dichos parámetros.

```
best_model<-glmnet(mx,my,keep=TRUE,alpha=0.7,lambda= 0.09708317)
```

Ejercicio 3

> **Note**
>
> Realizar un resumen del modelo obtenido, comparando los resultados con los obtenidos a través del modelo de regresión múltiple.

```
best_model$df
```

```
[1] 25
```

Se han seleccionado 25 variables. Veamos los valores estimados para cada β_i.

```
best_model$beta
```

```
36 x 1 sparse Matrix of class "dgCMatrix"
              s0
x01  1.082247085
x02  .
x03  1.091369327
x04  0.849497512
x05  0.976961984
x06  0.938065188
x07  .
x08  1.319208468
x10  0.774752835
x11  0.010848439
x13  0.032456761
x14  .
x15 -0.043701585
x16  0.888566187
x17  .
x18  .
x19  .
x20  .
x21 -0.070726161
x22 -0.003109018
x23  0.021867270
x25  .
x26  .
x27  0.101890957
x28  .
x29 -0.002521201
x30  0.070325491
x31 -0.011585123
x32  0.016469330
x33  0.088379251
x34 -0.007476030
x35 -0.048612149
x36  0.080660688
x37  .
x39  0.114783383
x40  0.018296150
```

Dos de las variables que presentaban una alta correlación eran X_{04} y X_{07}. Observamos que en el modelo de regularización se descarta la variable X_{07}. Otras dos variables que presentaban una alta correlación eran X_{10} y X_{11}. En esta ocasión se seleccionan ambas variables pero el estimador $\hat{\beta}_{11}$ toma un valor cercano a cero. Por tanto, los resultados obtenidos por el modelo de regularización son bastante coherentes, ya que casi no tiene en cuenta las variables que producen multicolinealidad.

Calculemos el coeficiente R^2 asociado a cada uno de los modelos y comparemos los resultados.

```
summary(regre)$r.squared
```

```
[1] 0.9137753
```

```
best_model$dev.ratio
```

```
[1] 0.9122242
```

El modelo de regresión lineal con todas las variables presenta un R^2 de 0.9137753, mientras que el modelo `elastic net`, considerando menos variables, obtiene un R^2 de 0.9122242. Podemos concluir por lo tanto, que en el modelo de regresión con regularización hemos podido penalizar a aquellas variables que presentaban problemas de multicolinealidad y que eran menos significativas con un decrecimiento del R^2 despreciable.