

Problema 1.

a) Análisis de Multicolinealidad

```
Call:
lm(formula = Apps ~ ., data = College1)

Residuals:
    Min       1Q   Median       3Q      Max
-8964    -776     -92     518   33186

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.605e+03  7.474e+02  -3.486  0.000519 ***
PrivateYes   -7.094e+02  2.547e+02  -2.785  0.005479 **
Top10perc    2.177e+01  1.020e+01   2.135  0.033110 *
Top25perc    -2.595e+00  8.241e+00  -0.315  0.752936
F.Undergrad   6.351e-01  2.210e-02  28.737 < 2e-16 ***
P.Undergrad  -9.153e-02  5.898e-02  -1.552  0.121069
Outstate     5.351e-02  3.472e-02   1.541  0.123732
Room.Board   3.621e-01  8.849e-02   4.092  4.72e-05 ***
Books        1.811e-01  4.408e-01   0.411  0.681306
Personal     -1.584e-01  1.164e-01  -1.361  0.173810
PhD          -1.503e+00  8.571e+00  -0.175  0.860889
Terminal     -1.059e+01  9.414e+00  -1.125  0.261090
S.F.Ratio    2.233e+01  2.405e+01   0.928  0.353483
perc.alumni  -2.469e+01  7.485e+00  -3.299  0.001016 **
Expend       1.010e-01  2.281e-02   4.429  1.09e-05 ***
Grad.Rate    2.438e+01  5.409e+00   4.508  7.58e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1926 on 761 degrees of freedom
Multiple R-squared:  0.7572,    Adjusted R-squared:  0.7524
F-statistic: 158.2 on 15 and 761 DF,  p-value: < 2.2e-16
```



Notemos que (Top10, top25) tienen una alta correlación y además tienen un VIF (Variance Inflation Factor) es mayor a 5, por tanto existe evidencia para posiblemente retirar alguna de las variables, restaría hacer un análisis más profundo (Opcional).

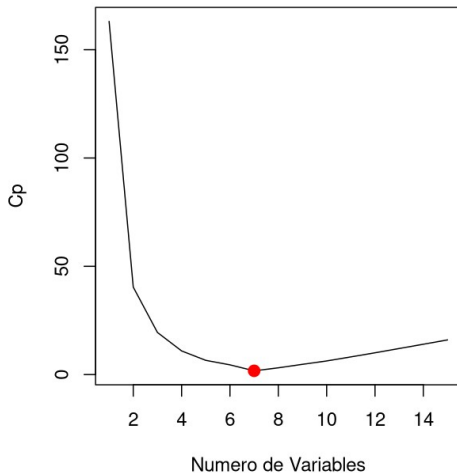
```
Private    Top10perc  Top25perc  F.Undergrad  P.Undergrad  Outstate  Room.Board
2.696029   6.772485   5.573798   2.403895     1.686748     4.083032   1.970365
Books      Personal    PhD       Terminal    S.F.Ratio  perc.alumni  Expend
1.108457   1.298778   4.098248   4.018677     1.896088     1.800194     2.968696
Grad.Rate
1.806293
```

b) Partición de los datos.

```
#Training
train = College1[c(1:600),]
#Test
test= College1[c(601:777),]
```

c) Modelos obtenidos:

- Mejor subconjunto (7 variables)



Call:

```
lm(formula = Apps ~ Private + Top10perc + F.Undergrad + Room.Board +
    perc.alumni + Expend + Grad.Rate, data = College1)
```

Residuals:

Min	1Q	Median	3Q	Max
-8572	-766	-93	537	33138

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-3.292e+03	3.529e+02	-9.328	< 2e-16 ***
Private	-4.846e+02	2.257e+02	-2.147	0.03207 *
Top10perc	1.883e+01	5.846e+00	3.220	0.00134 **
F.Undergrad	6.106e-01	1.932e-02	31.604	< 2e-16 ***
Room.Board	3.745e-01	7.824e-02	4.786	2.04e-06 ***
perc.alumni	-2.365e+01	7.177e+00	-3.295	0.00103 **
Expend	9.877e-02	1.929e-02	5.120	3.86e-07 ***
Grad.Rate	2.780e+01	5.198e+00	5.348	1.17e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1930 on 769 degrees of freedom

Multiple R-squared: 0.7535, Adjusted R-squared: 0.7512

F-statistic: 335.7 on 7 and 769 DF, p-value: < 2.2e-16

- RIDGE VS LASSO

15 x 1 sparse Matrix of class "dgCMatrix"

```
s0
Private      -762.83811387
Top10perc    13.26222406
Top25perc    5.11134011
F.Undergrad  0.61391496
P.Undergrad  0.08421817
Outstate     0.01053757
Room.Board   0.33115559
Books        0.35798337
Personal     -0.07588683
PhD          1.09697871
Terminal     -6.10919529
S.F.Ratio    8.29868307
perc.alumni  -18.21532249
Expend       0.12102191
Grad.Rate    19.08539402
```

> |

15 x 1 sparse Matrix of class "dgCMatrix"

```
s0
Private      -460.06925176
Top10perc    12.35666444
Top25perc    1.40498768
F.Undergrad  0.65089026
P.Undergrad  .
Outstate     .
Room.Board   0.32217772
Books        0.07008391
Personal     .
PhD          .
Terminal     .
S.F.Ratio    .
perc.alumni  -11.57753237
Expend       0.11125574
Grad.Rate    15.92425280
```

MSA (Inciso d)

Mejor subconjunto 3454699

Ridge 3430602

Lasso 3412933