Lista de ejercicios 3 - Análisis de regresión

Ciclo nivelación 2025-2

Mg. Sc. J. Eduardo Gamboa U.

Una empresa de análisis inmobiliario desea modelar el precio de venta de departamentos en una ciudad usando información estructural y comercial. Se construyó una base de datos de 100 propiedades vendidas recientemente.

Variable respuesta: Precio de venta (miles de USD)

Variables explicativas:

- x1: Área construida, en metros cuadrados
- x2: Índice de calidad constructiva (z-score), es un puntaje técnico de materiales y acabados.
- x3: Intensidad de publicidad digital (z-score), es una medida de exposición en portales y anuncios.
- x4: Tiene cochera (0: No, 1: Sí)
- x5: Zona de la ciudad (A = Zona periférica, B = Zona intermedia, C = Zona comercial, D = Zona premium)

Objetivos:

- Identificar leverages, residuales y valores influenciales
- Seleccionar variables
- Reportar el modelo resultante e interpretar sus coeficientes estimados.

1. División del conjunto de datos en entrenamiento y prueba

```r
library(readxl)
datos = read_xlsx('Lista3_datos.xlsx')
```

```r
library(rsample)
set.seed(12369874)
split_obj = initial_split(datos, prop = 0.75)
train     = training(split_obj)
test      = testing(split_obj)
train |> nrow()
```

```
[1] 375
```

```r
test |> nrow()
```

```
[1] 125
```

2. Creación del modelo inicial de regresión lineal

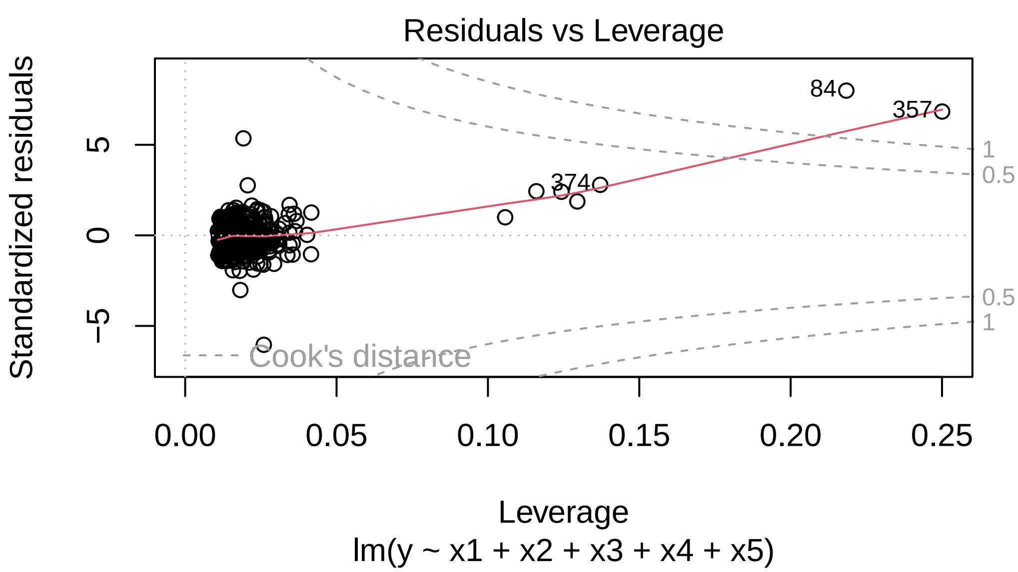
```r
modelo = lm(y ~ x1 + x2 + x3 + x4 + x5, data = train)
modelo |> summary()
```

```

Call:
lm(formula = y ~ x1 + x2 + x3 + x4 + x5, data = train)

Residuals:
     Min       1Q   Median       3Q      Max 
-11.7807  -1.0193  -0.1207   0.8529  13.9607 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  5.55013    0.42123  13.176  < 2e-16 ***
x1           1.35969    0.04177  32.553  < 2e-16 ***
x2          -0.55100    0.06354  -8.672  < 2e-16 ***
x3           0.02829    0.10227   0.277 0.782246    
x41          0.98436    0.20580   4.783 2.51e-06 ***
x5B          0.77653    0.27400   2.834 0.004851 ** 
x5C         -1.02727    0.26210  -3.919 0.000106 ***
x5D          1.18452    0.32541   3.640 0.000312 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.976 on 367 degrees of freedom
Multiple R-squared:  0.9966,	Adjusted R-squared:  0.9965 
F-statistic: 1.532e+04 on 7 and 367 DF,  p-value: < 2.2e-16
```

3. Detección de leverages

```
modelo |> model.matrix() -> X
X %*% solve(t(X) %*% X) %*% t(X) -> H
train |> nrow() -> n
modelo |> coef() |> length() -> k
as.vector(which(diag(H) > 2*k/n)) -> leverages
```

```
library(olsrr)
modelo |> plot(which=5)
```



Leverages identificados:

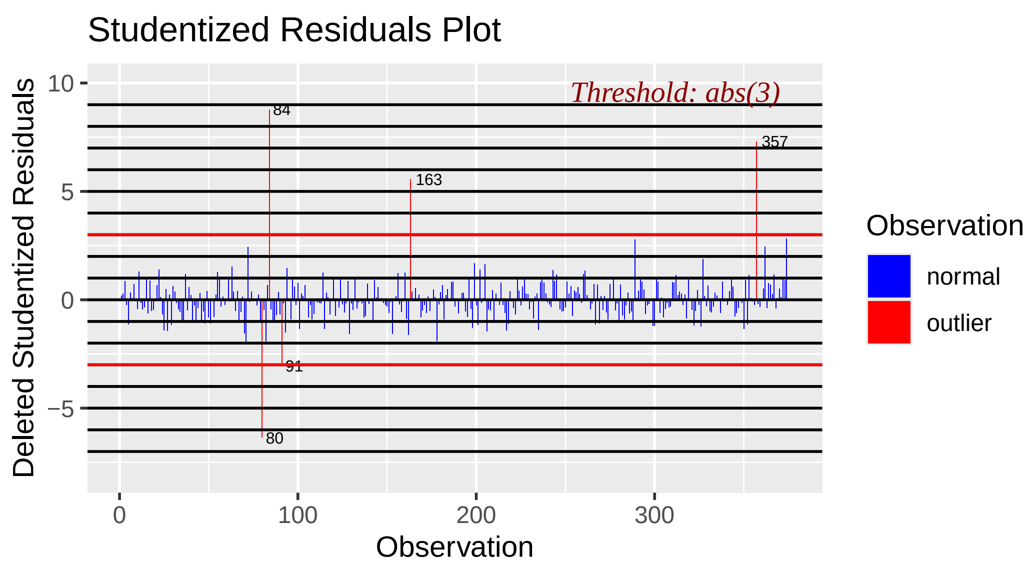
```
leverages
```

```
[1]  72  84 319 327 357 362 374
```

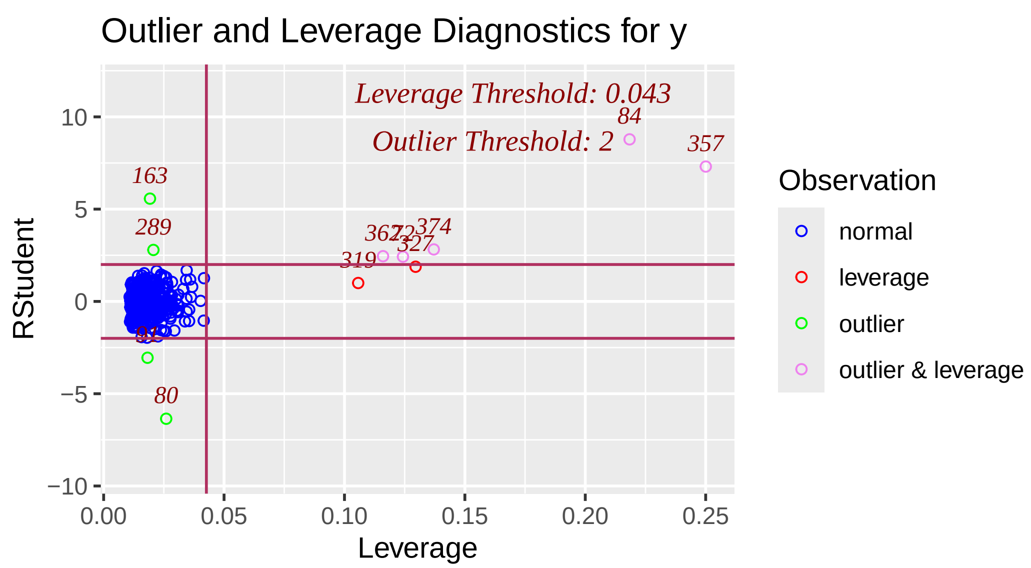
4. Detección de valores atípicos

```
library(MASS)
as.vector(which((modelo |> studres() |> abs()) > 2)) -> outliers
```

```
modelo |> ols_plot_resid_stud()
```




```
modelo |> ols_plot_resid_lev()
```



Valores atípicos identificados:

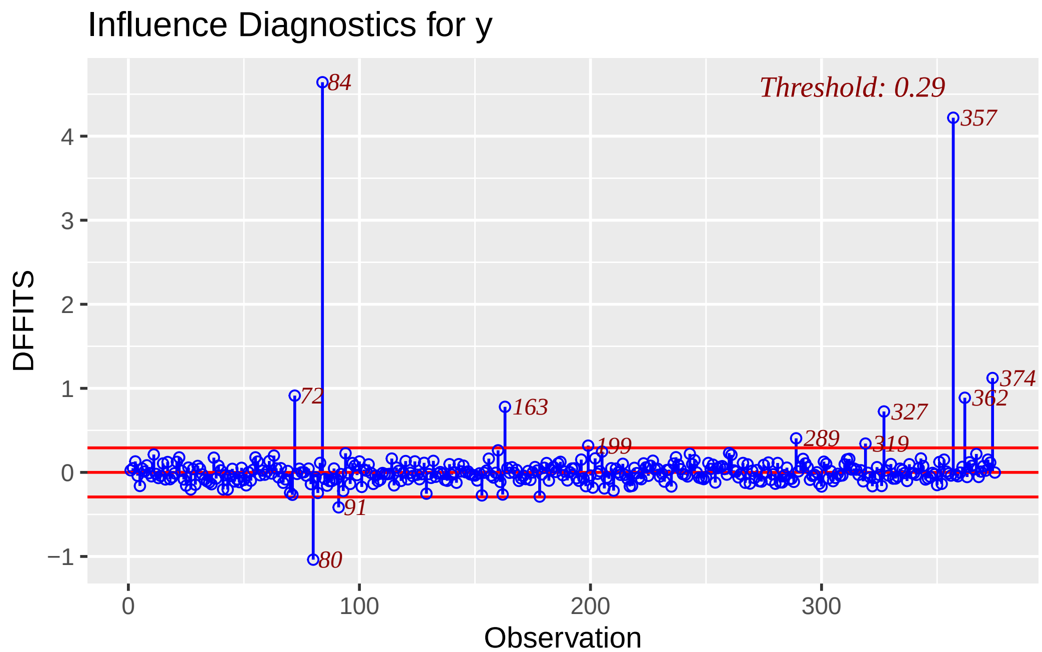
```r
outliers
```

```
[1]  72  80  84  91 163 289 357 362 374
```

5. Detección de DFFITS

```
modelo |> coef() |> length() -> k
train |> nrow() -> n
(abs(modelo |> dffits())) -> dffits
as.vector(which(dffits >= 2*sqrt(k/n))) -> inf_dffits
```

```
modelo |> ols_plot_dffits()
```



Valores influyentes según DFFITS:

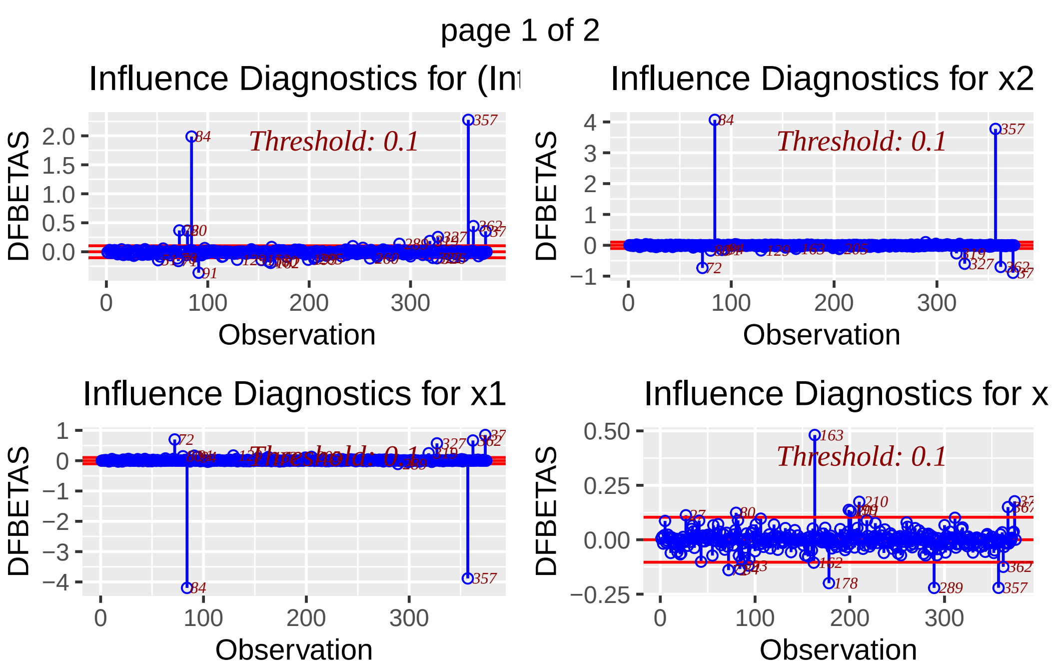
```
inf_dffits
```

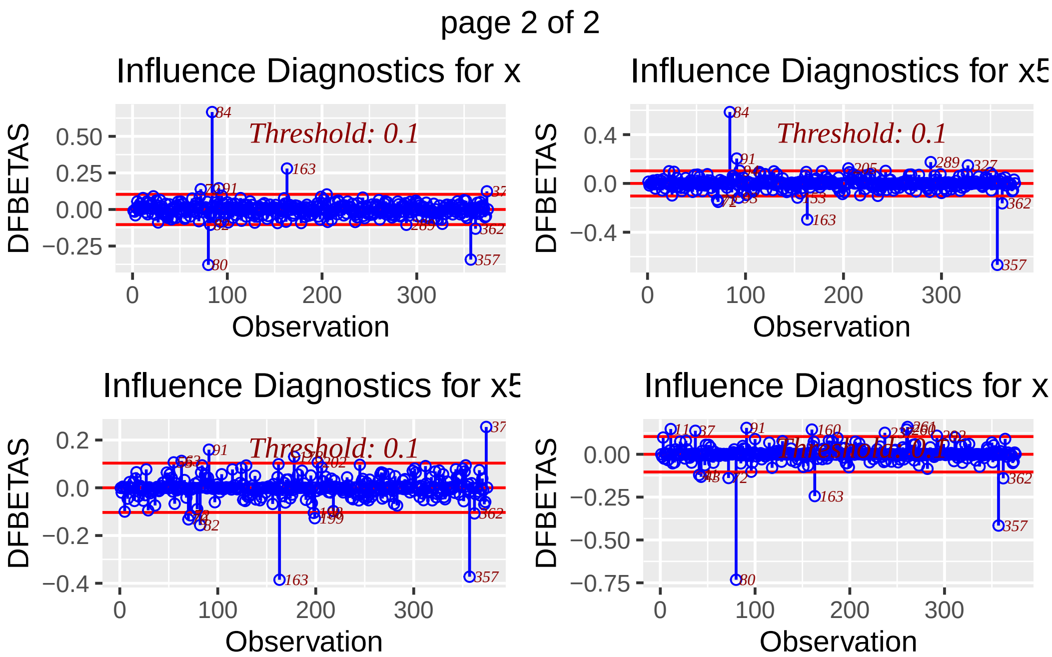
```
[1]  72  80  84  91 163 199 289 319 327 357 362 374
```

6. Detección de DFBETAS

```r
(abs(modelo |> dfbetas())) -> dfbetas
as.vector(which(dfbetas[,1] >= 2/sqrt(n))) -> dfbeta0

as.vector(which(dfbetas[,2] >= 2/sqrt(n))) -> dfbeta1

as.vector(which(dfbetas[,3] >= 2/sqrt(n))) -> dfbeta2

as.vector(which(dfbetas[,4] >= 2/sqrt(n))) -> dfbeta3

as.vector(which(dfbetas[,5] >= 2/sqrt(n))) -> dfbeta4

as.vector(which(dfbetas[,6] >= 2/sqrt(n))) -> dfbeta5

as.vector(which(dfbetas[,7] >= 2/sqrt(n))) -> dfbeta6

as.vector(which(dfbetas[,8] >= 2/sqrt(n))) -> dfbeta7
```

```r
modelo |> ols_plot_dfbetas()
```





Valores influyentes según DFBETAS:

```
dfbeta0
```

```
 [1]  51  70  71  72  80  84  91 129 153 160 162 199 205 260 289 319 322 326 327
[20] 357 362 374
```

```
dfbeta1
```

```
 [1]  72  80  84  91  94 129 163 205 289 319 327 357 362 374
```

```
dfbeta2
```

```
 [1]  72  80  84  91  94 129 163 205 319 327 357 362 374
```

```
dfbeta3
```

```
 [1]  27  72  80  84  93 162 163 178 199 201 210 289 357 362 367 374
```

```
dfbeta4
```

```
 [1]  72  80  82  84  91 163 289 357 362 374
```

```
dfbeta5
```

```
 [1]  55  63  70  72  82  91 163 178 198 199 202 357 362 374
```

```
dfbeta6
```

```
 [1]  71  72  84  91  93  94 153 163 205 289 327 357 362
```

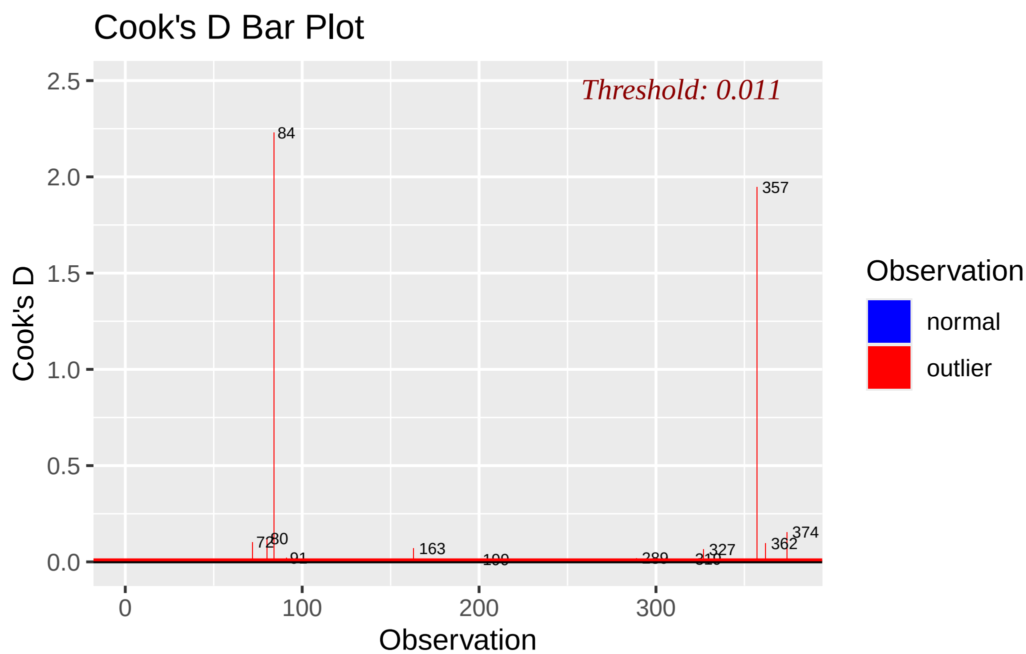
```
dfbeta7
```

```
 [1]  11  37  41  43  72  80  91 160 163 237 260 261 292 357 362
```

7. Detección de valores influyentes según distancia de Cook

```
modelo |> cooks.distance() -> cookd
as.vector(which(cookd > 4/n)) -> inf_cook
```

```
modelo |> ols_plot_cooksd_bar()
```



Valores influyentes según Distancia de Cook:

```
inf_cook
```

```
 [1]  72  80  84  91 163 199 289 319 327 357 362 374
```

8. Detección de valores influyentes según COVRATIO

```
modelo |> covratio() |> abs() -> covra
as.vector(which(covra > 1+3*k/n)) -> inf_covra1
as.vector(which(covra < 1-3*k/n)) -> inf_covra2
c(inf_covra1, inf_covra2) -> inf_covra
```

Valores influyentes según COVRATIO:

```
inf_covra
```

```
[1] 185 319 327  80  84  91 163 289 357
```

En resumen:

```
lista <- list(leverages,
              outliers,
              inf_dffits,
              dfbeta0,dfbeta1,dfbeta2,dfbeta3,dfbeta4, dfbeta5,dfbeta6,dfbeta7,
              inf_cook,
              inf_covra)

inter <- Reduce(intersect, lista)
inter
```

```
[1] 357
```

9. Remoción de los valores leverage, atípicos e influyentes

```
train2 <- train[-c(inter),]
```

10. Creación de un nuevo modelo

```
modelo2 = lm(y ~ x1 + x2 + x3 + x4 + x5, data = train2)
```

11. Comparación de los reportes del modelo original y el nuevo

```r
modelo |> summary()
```

```

Call:
lm(formula = y ~ x1 + x2 + x3 + x4 + x5, data = train)

Residuals:
     Min       1Q   Median       3Q      Max 
-11.7807  -1.0193  -0.1207   0.8529  13.9607 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  5.55013    0.42123  13.176  < 2e-16 ***
x1           1.35969    0.04177  32.553  < 2e-16 ***
x2          -0.55100    0.06354  -8.672  < 2e-16 ***
x3           0.02829    0.10227   0.277 0.782246    
x41          0.98436    0.20580   4.783 2.51e-06 ***
x5B          0.77653    0.27400   2.834 0.004851 ** 
x5C         -1.02727    0.26210  -3.919 0.000106 ***
x5D          1.18452    0.32541   3.640 0.000312 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.976 on 367 degrees of freedom
Multiple R-squared:  0.9966,	Adjusted R-squared:  0.9965 
F-statistic: 1.532e+04 on 7 and 367 DF,  p-value: < 2.2e-16
```

```r
modelo2 |> summary()
```

```

Call:
lm(formula = y ~ x1 + x2 + x3 + x4 + x5, data = train2)

Residuals:
     Min       1Q   Median       3Q      Max 
-11.7913  -0.9513  -0.1185   0.8672  17.3718 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.65340    0.41271  11.275  < 2e-16 ***
x1           1.51145    0.04425  34.158  < 2e-16 ***
x2          -0.77528    0.06690 -11.589  < 2e-16 ***
x3           0.04942    0.09571   0.516 0.605908    
x41          1.05033    0.19273   5.450 9.29e-08 ***
x5B          0.87202    0.25665   3.398 0.000754 ***
x5C         -0.86361    0.24620  -3.508 0.000508 ***
x5D          1.31120    0.30490   4.300 2.19e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.848 on 366 degrees of freedom
Multiple R-squared:  0.9969,	Adjusted R-squared:  0.9969 
F-statistic: 1.688e+04 on 7 and 366 DF,  p-value: < 2.2e-16
```

12. Selección de variables según R^2 predictivo

```
library(olsrr)
library(dplyr)
ols_step_all_possible(modelo2)$result |> data.frame() -> resultados
resultados |>
  select(n,predrsq,predictors) |>
  arrange(-predrsq) |>
  head()
```

```
  n   predrsq     predictors
1 4 0.9960979    x1 x2 x4 x5
2 5 0.9960822 x1 x2 x3 x4 x5
3 3 0.9958170       x1 x2 x5
4 4 0.9958046    x1 x2 x3 x5
5 3 0.9956196       x1 x2 x4
6 4 0.9956054    x1 x2 x3 x4
```

Modelo resultante: Y = f(x1, x5)

13. Selección de variables según AIC

```
resultados |>
  select(n,aic,predictors) |>
  arrange(aic) |>
  head()
```

```
  n      aic     predictors
1 4 1529.078    x1 x2 x4 x5
2 5 1530.806 x1 x2 x3 x4 x5
3 3 1556.697       x1 x2 x5
4 4 1557.986    x1 x2 x3 x5
5 3 1586.814       x1 x2 x4
6 4 1587.844    x1 x2 x3 x4
```

Modelo resultante: Y = f(x1, x2, x4, x5)

14. Selección de variables según SBC

```
resultados |>
  select(n,sbc,predictors) |>
  arrange(sbc) |>
  head()
```

```
  n      sbc      predictors
1 4 1560.472    x1 x2 x4 x5
2 5 1566.124 x1 x2 x3 x4 x5
3 3 1584.167       x1 x2 x5
4 4 1589.380    x1 x2 x3 x5
5 3 1606.435       x1 x2 x4
6 4 1611.390    x1 x2 x3 x4
```

Modelo resultante: Y = f(x1, x2, x4, x5)

15. Selección de variables según SBIC

```
resultados |>
  select(n,sbic,predictors) |>
  arrange(sbic) |>
  head()
```

```
  n     sbic      predictors
1 4 463.9225    x1 x2 x4 x5
2 5 465.7012 x1 x2 x3 x4 x5
3 3 490.8719       x1 x2 x5
4 4 492.0656    x1 x2 x3 x5
5 3 524.2984       x1 x2 x4
6 4 525.0742    x1 x2 x3 x4
```

Modelo resultante: Y = f(x1, x2, x4, x5)

16. Selección de variables según Cp de Mallows

```r
resultados |>
  select(n,cp,predictors) |>
  mutate(dif = n-cp,.after=cp) |>
  arrange(abs(dif)) |>
  head()
```

```
  n        cp        dif     predictors
1 4  6.266637  -2.266637    x1 x2 x4 x5
2 5  8.000000  -3.000000 x1 x2 x3 x4 x5
3 3 34.452883 -31.452883       x1 x2 x5
4 4 35.699554 -31.699554    x1 x2 x3 x5
5 4 69.193442 -65.193442    x1 x2 x3 x4
6 3 68.318519 -65.318519       x1 x2 x4
```

Modelo resultante: Y = f(X1, X2, X4, X5)

17. Selección del modelo según mejores subconjuntos

```r
modelo2 |> ols_step_best_subset()
```

```
   Best Subsets Regression
-----------------------------
Model Index    Predictors
-----------------------------
     1         x1
     2         x1 x2
     3         x1 x2 x5
     4         x1 x2 x4 x5
     5         x1 x2 x3 x4 x5
-----------------------------

                                                      Subsets Regression Summary
-------------------------------------------------------------------------------------------------------------------------------
                       Adj.        Pred
Model    R-Square    R-Square    R-Square      C(p)          AIC          SBIC          SBC          MSEP         FPE       HSP
-------------------------------------------------------------------------------------------------------------------------------
  1        0.9949      0.9948      0.9947    239.6075    1709.6139    646.4032    1721.3867    2094.0560    5.6290    0.0151
  2        0.9960      0.9960      0.9953    102.0304    1614.3672    551.6919    1630.0642    1618.9601    4.3635    0.0117
  3        0.9967      0.9966      0.9958     34.4529    1556.6974    490.8719    1584.1672    1369.2324    3.7203    0.0100
  4        0.9969      0.9969      0.9961      6.2666    1529.0782    463.9225    1560.4723    1268.4153    3.4555    0.0093
  5        0.9969      0.9969      0.9961      8.0000    1530.8059    465.7012    1566.1242    1270.9456    3.4715    0.0093
-------------------------------------------------------------------------------------------------------------------------------
AIC: Akaike Information Criteria
 SBIC: Sawa's Bayesian Information Criteria
 SBC: Schwarz Bayesian Criteria
 MSEP: Estimated error of prediction, assuming multivariate normality
 FPE: Final Prediction Error
 HSP: Hocking's Sp
 APC: Amemiya Prediction Criteria
```

Modelo resultante: Y = f()

18. Selección del mejor modelo según forward selection

```
modelo2 |> ols_step_forward_p()
```

```
                                Stepwise Summary
-----------------------------------------------------------------------------
Step    Variable        AIC         SBC         SBIC        R2       Adj. R2
-----------------------------------------------------------------------------
 0      Base Model    3678.527    3686.375    2613.180    0.00000    0.00000
 1      x1            1709.614    1721.387     646.403    0.99486    0.99484
 2      x2            1614.367    1630.064     551.692    0.99603    0.99601
 3      x4            1586.814    1606.435     524.298    0.99633    0.99631
 4      x5            1529.078    1560.472     463.922    0.99691    0.99686
-----------------------------------------------------------------------------

Final Model Output
------------------

                          Model Summary
-----------------------------------------------------------------
R                         0.998       RMSE                 1.829
R-Squared                 0.997       MSE                  3.346
Adj. R-Squared            0.997       Coef. Var            1.456
Pred R-Squared            0.996       AIC               1529.078
MAE                       1.178       SBC               1560.472
-----------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error
 AIC: Akaike Information Criteria
 SBC: Schwarz Bayesian Criteria

                                  ANOVA
---------------------------------------------------------------------------
                  Sum of
                 Squares          DF    Mean Square        F          Sig.
---------------------------------------------------------------------------
Regression    403616.821           6      67269.470    19727.742    0.0000
Residual        1251.430         367          3.410
Total         404868.252         373
---------------------------------------------------------------------------

                                   Parameter Estimates
-----------------------------------------------------------------------------------------
      model       Beta    Std. Error    Std. Beta         t        Sig      lower    upper
-----------------------------------------------------------------------------------------
(Intercept)      4.646         0.412                   11.275     0.000     3.835    5.456
         x1      1.511         0.044        1.506      34.192     0.000     1.425    1.598
         x2     -0.775         0.067       -0.511     -11.600     0.000    -0.907   -0.644
        x41      1.057         0.192        0.016       5.500     0.000     0.679    1.435
        x5B      0.883         0.256        0.012       3.454     0.001     0.380    1.385
        x5C     -0.872         0.245       -0.012      -3.554     0.000    -1.355   -0.390
        x5D      1.300         0.304        0.014       4.279     0.000     0.703    1.898
-----------------------------------------------------------------------------------------
```

```
modelo2 |> ols_step_forward_aic()
```

Stepwise Summary

Step	Variable	AIC	SBC	SBIC	R2	Adj. R2
0	Base Model	3678.527	3686.375	2613.180	0.00000	0.00000
1	x1	1709.614	1721.387	646.403	0.99486	0.99484
2	x2	1614.367	1630.064	551.692	0.99603	0.99601
3	x5	1556.697	1584.167	490.872	0.99665	0.99661
4	x4	1529.078	1560.472	463.922	0.99691	0.99686

Final Model Output

Model Summary

R	0.998	RMSE	1.829
R-Squared	0.997	MSE	3.346
Adj. R-Squared	0.997	Coef. Var	1.456
Pred R-Squared	0.996	AIC	1529.078
MAE	1.178	SBC	1560.472

RMSE: Root Mean Square Error
MSE: Mean Square Error
MAE: Mean Absolute Error
AIC: Akaike Information Criteria
SBC: Schwarz Bayesian Criteria

ANOVA

	Sum of Squares	DF	Mean Square	F	Sig.
Regression	403616.821	6	67269.470	19727.742	0.0000
Residual	1251.430	367	3.410		
Total	404868.252	373			

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	4.646	0.412		11.275	0.000	3.835	5.456
x1	1.511	0.044	1.506	34.192	0.000	1.425	1.598
x2	-0.775	0.067	-0.511	-11.600	0.000	-0.907	-0.644
x5B	0.883	0.256	0.012	3.454	0.001	0.380	1.385
x5C	-0.872	0.245	-0.012	-3.554	0.000	-1.355	-0.390
x5D	1.300	0.304	0.014	4.279	0.000	0.703	1.898
x41	1.057	0.192	0.016	5.500	0.000	0.679	1.435

Modelo resultante: Y = f()

19. Selección del mejor modelo según backward selection

```r
modelo2 |> ols_step_backward_p()
```

```
                             Stepwise Summary
----------------------------------------------------------------------------
Step    Variable        AIC          SBC         SBIC        R2       Adj. R2
----------------------------------------------------------------------------
 0      Full Model    1530.806     1566.124     465.701    0.99691    0.99685
 1      x3            1529.078     1560.472     463.922    0.99691    0.99686
----------------------------------------------------------------------------

Final Model Output
------------------

                        Model Summary
----------------------------------------------------------------
R                         0.998       RMSE                1.829
R-Squared                 0.997       MSE                 3.346
Adj. R-Squared            0.997       Coef. Var           1.456
Pred R-Squared            0.996       AIC              1529.078
MAE                       1.178       SBC              1560.472
----------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error
 AIC: Akaike Information Criteria
 SBC: Schwarz Bayesian Criteria

                                 ANOVA
---------------------------------------------------------------------------
                  Sum of
                 Squares         DF    Mean Square         F          Sig.
---------------------------------------------------------------------------
Regression    403616.821          6      67269.470     19727.742    0.0000
Residual        1251.430        367          3.410
Total         404868.252        373
---------------------------------------------------------------------------

                                  Parameter Estimates
-------------------------------------------------------------------------------------------
      model       Beta    Std. Error    Std. Beta        t        Sig       lower     upper
-------------------------------------------------------------------------------------------
(Intercept)      4.646         0.412                  11.275     0.000      3.835     5.456
         x1      1.511         0.044        1.506     34.192     0.000      1.425     1.598
         x2     -0.775         0.067       -0.511    -11.600     0.000     -0.907    -0.644
        x41      1.057         0.192        0.016      5.500     0.000      0.679     1.435
        x5B      0.883         0.256        0.012      3.454     0.001      0.380     1.385
        x5C     -0.872         0.245       -0.012     -3.554     0.000     -1.355    -0.390
        x5D      1.300         0.304        0.014      4.279     0.000      0.703     1.898
-------------------------------------------------------------------------------------------
```

```r
modelo2 |> ols_step_backward_aic()
```

```
                             Stepwise Summary
```

Step	Variable	AIC	SBC	SBIC	R2	Adj. R2
0	Full Model	1530.806	1566.124	465.701	0.99691	0.99685
1	x3	1529.078	1560.472	463.902	0.99691	0.99686

Final Model Output

Model Summary

R	0.998	RMSE	1.829
R-Squared	0.997	MSE	3.346
Adj. R-Squared	0.997	Coef. Var	1.456
Pred R-Squared	0.996	AIC	1529.078
MAE	1.178	SBC	1560.472

RMSE: Root Mean Square Error
MSE: Mean Square Error
MAE: Mean Absolute Error
AIC: Akaike Information Criteria
SBC: Schwarz Bayesian Criteria

ANOVA

	Sum of Squares	DF	Mean Square	F	Sig.
Regression	403616.821	6	67269.470	19727.742	0.0000
Residual	1251.430	367	3.410		
Total	404868.252	373			

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	4.646	0.412		11.275	0.000	3.835	5.456
x1	1.511	0.044	1.506	34.192	0.000	1.425	1.598
x2	-0.775	0.067	-0.511	-11.600	0.000	-0.907	-0.644
x41	1.057	0.192	0.016	5.500	0.000	0.679	1.435
x5B	0.883	0.256	0.012	3.454	0.001	0.380	1.385
x5C	-0.872	0.245	-0.012	-3.554	0.000	-1.355	-0.390
x5D	1.300	0.304	0.014	4.279	0.000	0.703	1.898

Modelo resultante: Y = f()

20. Selección del mejor modelo según stepwise selection

```r
modelo2 |> ols_step_both_p()
```

```
                              Stepwise Summary
-----------------------------------------------------------------------------
Step    Variable        AIC         SBC         SBIC        R2       Adj. R2
-----------------------------------------------------------------------------
 0      Base Model    3678.527    3686.375    2613.180    0.00000    0.00000
 1      x1 (+)        1709.614    1721.387     646.403    0.99486    0.99484
 2      x2 (+)        1614.367    1630.064     551.692    0.99603    0.99601
 3      x4 (+)        1586.814    1606.435     524.298    0.99633    0.99631
 4      x5 (+)        1529.078    1560.472     463.922    0.99691    0.99686
-----------------------------------------------------------------------------

Final Model Output
------------------

                         Model Summary
----------------------------------------------------------------
R                        0.998       RMSE                 1.829
R-Squared                0.997       MSE                  3.346
Adj. R-Squared           0.997       Coef. Var            1.456
Pred R-Squared           0.996       AIC               1529.078
MAE                      1.178       SBC               1560.472
----------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error
 AIC: Akaike Information Criteria
 SBC: Schwarz Bayesian Criteria

                                  ANOVA
--------------------------------------------------------------------------
                 Sum of
                Squares         DF    Mean Square        F          Sig.
--------------------------------------------------------------------------
Regression    403616.821         6      67269.470    19727.742    0.0000
Residual        1251.430       367          3.410
Total         404868.252       373
--------------------------------------------------------------------------

                                   Parameter Estimates
---------------------------------------------------------------------------------------
      model       Beta    Std. Error    Std. Beta         t        Sig      lower    upper
---------------------------------------------------------------------------------------
(Intercept)      4.646         0.412                  11.275     0.000     3.835    5.456
         x1      1.511         0.044        1.506     34.192     0.000     1.425    1.598
         x2     -0.775         0.067       -0.511    -11.600     0.000    -0.907   -0.644
        x41      1.057         0.192        0.016      5.500     0.000     0.679    1.435
        x5B      0.883         0.256        0.012      3.454     0.001     0.380    1.385
        x5C     -0.872         0.245       -0.012     -3.554     0.000    -1.355   -0.390
        x5D      1.300         0.304        0.014      4.279     0.000     0.703    1.898
---------------------------------------------------------------------------------------
```

```r
modelo2 |> ols_step_both_aic()
```

Stepwise Summary

Step	Variable	AIC	SBC	SBIC	R2	Adj. R2
0	Base Model	3678.527	3686.375	2613.180	0.00000	0.00000
1	x1 (+)	1709.614	1721.387	646.403	0.99486	0.99484
2	x2 (+)	1614.367	1630.064	551.692	0.99603	0.99601
3	x5 (+)	1556.697	1584.167	490.872	0.99665	0.99661
4	x4 (+)	1529.078	1560.472	463.922	0.99691	0.99686

Final Model Output

Model Summary

R	0.998	RMSE	1.829
R-Squared	0.997	MSE	3.346
Adj. R-Squared	0.997	Coef. Var	1.456
Pred R-Squared	0.996	AIC	1529.078
MAE	1.178	SBC	1560.472

RMSE: Root Mean Square Error
MSE: Mean Square Error
MAE: Mean Absolute Error
AIC: Akaike Information Criteria
SBC: Schwarz Bayesian Criteria

ANOVA

	Sum of Squares	DF	Mean Square	F	Sig.
Regression	403616.821	6	67269.470	19727.742	0.0000
Residual	1251.430	367	3.410		
Total	404868.252	373			

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	4.646	0.412		11.275	0.000	3.835	5.456
x1	1.511	0.044	1.506	34.192	0.000	1.425	1.598
x2	-0.775	0.067	-0.511	-11.600	0.000	-0.907	-0.644
x5B	0.883	0.256	0.012	3.454	0.001	0.380	1.385
x5C	-0.872	0.245	-0.012	-3.554	0.000	-1.355	-0.390
x5D	1.300	0.304	0.014	4.279	0.000	0.703	1.898
x41	1.057	0.192	0.016	5.500	0.000	0.679	1.435

Modelo resultante: Y = f()

21. Uso del testing

```r
pred_train <- predict(modelo2, newdata = train2)
pred_test  <- predict(modelo2, newdata = test)
```

```r
metricas <- function(y, yhat){
  rmse <- sqrt(mean((y - yhat)^2))
  mae  <- mean(abs(y - yhat))
  r2   <- 1 - sum((y - yhat)^2)/sum((y - mean(y))^2)
  c(RMSE = rmse, MAE = mae, R2 = r2)
}
```

```r
m_train <- metricas(train2$y, pred_train)
m_test  <- metricas(test$y,  pred_test)

m_train
```

```
     RMSE       MAE        R2 
1.8285608 1.1797236 0.9969113 
```

```r
m_test
```

```
     RMSE       MAE        R2 
2.6046388 1.3963764 0.9931583 
```

21. Interpretación de coeficientes del mejor modelo