$$Y \sim X_1 + X_2 + X_3 + X_4$$

$$\boxed{\uparrow X_1 \quad \uparrow Y} \Rightarrow \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4$$

$$t_{calc} = \frac{\hat{\beta} - \beta_0}{s_\beta} \downarrow$$

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -54.28225    7.39453  -7.341 5.41e-08 ***
aptitude      0.32356    0.06778   4.774 5.15e-05 ***
tol           0.03337    0.07124   0.468   0.6431    
technical     1.09547    0.18138   6.039 1.65e-06 ***
general       0.53683    0.15840   3.389   0.0021 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.549 on 28 degrees of freedom
Multiple R-squared:  0.8768,	Adjusted R-squared:  0.8592 
F-statistic: 49.81 on 4 and 28 DF,  p-value: 2.467e-12
```

Factor de inflación de varianza (VIF): El VIF de la j-ésima variable viene dado por el j-ésimo término de la diagonal de $(\mathbf{X}'\mathbf{X})^{-1}$. También se puede hallar como $VIF_j = \frac{1}{1-R_j^2}$ donde R_j^2 es el coeficiente de determinación usando la j-ésima variable predictora como respuesta, y las demás predictoras como independientes. Su valor debe ser cercano a 1. Si es mayor a 5 es indicador de multicolinealidad. Si es superior a 10, esta multicolinealidad. es muy fuerte. Note la relación entre R^2 y VIF.

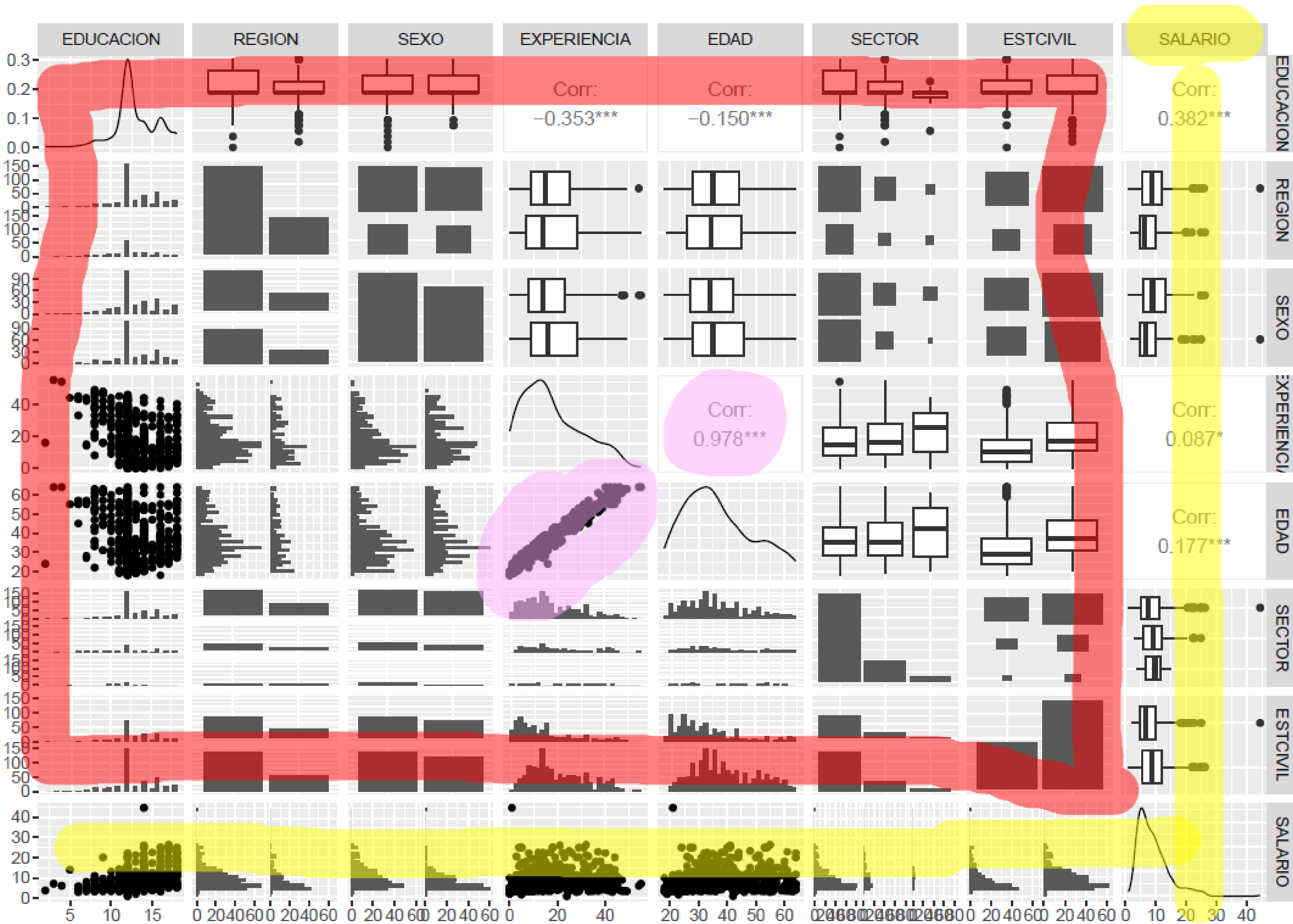
$$Y \sim X_1 + X_2 + X_3 + X_4$$

$VIF_1 = \frac{1}{1-R_1^2}$ donde R_1^2 es el coef. de determinación de $X_1 \sim X_2 + X_3 + X_4$

Ideal: $R_1^2 \Rightarrow$ bajo $\Rightarrow VIF \to 1$

$VIF_2 = \frac{1}{1-R_2^2}$ donde R_2^2 es el coef. de determinación de $X_2 \sim X_1 + X_3 + X_4$

Ideal: $R_2^2 \Rightarrow$ bajo



EDUCACION
REGION
SEXO
EXPERIENCIA
EDAD
SECTOR
ESTCIVIL
SALARIO
Corr:
-0.353***
Corr:
-0.150***
Corr:
0.382***
Corr:
0.978***
Corr:
0.087*
Corr:
0.177***

```
> modelo |> ols_step_backward_p(prem = 0.10)
```

```
                              Stepwise Summary
---------------------------------------------------------------------------
Step    Variable        AIC         SBC         SBIC         R2       Adj. R2
---------------------------------------------------------------------------
 0      Full Model    3116.178    3158.982    1599.025    0.26778    0.25662
 1      EDAD          3114.310    3152.833    1597.119    0.26760    0.25785
---------------------------------------------------------------------------

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

                         Model Summary
---------------------------------------------------------------
R                         0.517       RMSE                4.394
R-Squared                 0.268       MSE                19.307
Adj. R-Squared            0.258       Coef. Var          49.060
Pred R-Squared            0.246       AIC              3114.310
MAE                       3.143       SBC              3152.833
---------------------------------------------------------------
 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     3766.888         7        538.127    27.455    0.0000
Residual      10309.811       526         19.600
Total         14076.699       533
----------------------------------------------------------------------

                              Parameter Estimates
----------------------------------------------------------------------------------------
      model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
----------------------------------------------------------------------------------------
(Intercept)    -4.249         1.257                 -3.381    0.001    -6.717    -1.780
  EDUCACION     0.932         0.081        0.474    11.568    0.000     0.774     1.090
    REGION1    -0.839         0.428       -0.074    -1.959    0.051    -1.679     0.002
      SEXO1    -2.250         0.393       -0.218    -5.727    0.000    -3.022    -1.478
EXPERIENCIA     0.104         0.017        0.250     5.982    0.000     0.070     0.138
    SECTOR1     1.024         0.504        0.078     2.032    0.043     0.034     2.014
    SECTOR2     0.614         0.957        0.025     0.642    0.521    -1.266     2.495
  ESTCIVIL1     0.531         0.420        0.049     1.264    0.207    -0.294     1.356
----------------------------------------------------------------------------------------
```

modelo1 <- lm(SALARIO ~ EDUCACION + REGION + SEXO + EXPERIENCIA + SECTOR + ESTADO CIVIL)

```
> modelo |> ols_step_backward_aic()

                              Stepwise Summary
-----------------------------------------------------------------------------
Step    Variable        AIC         SBC         SBIC        R2       Adj. R2
-----------------------------------------------------------------------------
 0      Full Model    3116.178    3158.982    1599.025    0.26778    0.25662
 1      EDAD          3114.310    3152.833    1597.096    0.26760    0.25785
 2      ESTCIVIL      3113.929    3148.172    1596.662    0.26537    0.25701
-----------------------------------------------------------------------------

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

                        Model Summary
-----------------------------------------------------------------
R                       0.515       RMSE                   4.401
R-Squared               0.265       MSE                   19.365
Adj. R-Squared          0.257       Coef. Var             49.088
Pred R-Squared          0.247       AIC                 3113.929
MAE                     3.154       SBC                 3148.172
-----------------------------------------------------------------
 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    3735.570         6        622.595    31.728    0.0000
Residual     10341.129       527         19.623
Total        14076.699       533
-----------------------------------------------------------------------

                                     Parameter Estimates
-----------------------------------------------------------------------------------------
      model       Beta    Std. Error    Std. Beta      t        Sig      lower     upper
-----------------------------------------------------------------------------------------
(Intercept)     -4.113         1.253                -3.283    0.001    -6.574    -1.652
  EDUCACION      0.940         0.080        0.478   11.687    0.000     0.782     1.098
    REGION1     -0.828         0.428       -0.073   -1.933    0.054    -1.669     0.013
      SEXO1     -2.252         0.393       -0.219   -5.730    0.000    -3.025    -1.480
EXPERIENCIA      0.110         0.017        0.264    6.578    0.000     0.077     0.143
    SECTOR1      1.041         0.504        0.079    2.066    0.039     0.051     2.032
    SECTOR2      0.650         0.957        0.026    0.679    0.497    -1.230     2.531
-----------------------------------------------------------------------------------------
```

modelo2 <- lm(SALARIO ~ EDUCACION + REGION + SEXO + EXPERIENCIA + SECTOR)

```
> modelo |> ols_step_forward_p(prem = 0.10)

                                  Stepwise Summary
------------------------------------------------------------------------------
Step    Variable         AIC         SBC         SBIC        R2        Adj. R2
------------------------------------------------------------------------------
 0      Base Model     3266.610    3275.171    1750.543    0.00000    0.00000
 1      EDUCACION      3184.417    3197.258    1668.446    0.14586    0.14426
 2      EXPERIENCIA    3150.099    3167.220    1634.263    0.20202    0.19902
 3      SEXO           3116.735    3138.137    1601.287    0.25316    0.24893
 4      REGION         3114.403    3140.086    1599.030    0.25919    0.25359
 5      SECTOR         3113.929    3148.172    1596.668    0.26537    0.25701
 6      ESTCIVIL       3114.310    3152.833    1597.119    0.26760    0.25785
------------------------------------------------------------------------------

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

                            Model Summary
-----------------------------------------------------------------
R                         0.517       RMSE                 4.394
R-Squared                 0.268       MSE                 19.307
Adj. R-Squared            0.258       Coef. Var           49.060
Pred R-Squared            0.246       AIC               3114.310
MAE                       3.143       SBC               3152.833
-----------------------------------------------------------------
 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     3766.888         7        538.127    27.455    0.0000
Residual      10309.811       526         19.600
Total         14076.699       533
-----------------------------------------------------------------------

                                    Parameter Estimates
----------------------------------------------------------------------------------------
      model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
----------------------------------------------------------------------------------------
(Intercept)    -4.249         1.257                 -3.381    0.001    -6.717    -1.780
  EDUCACION     0.932         0.081        0.474    11.568    0.000     0.774     1.090
EXPERIENCIA     0.104         0.017        0.250     5.982    0.000     0.070     0.138
      SEXO1    -2.250         0.393       -0.218    -5.727    0.000    -3.022    -1.478
    REGION1    -0.839         0.428       -0.074    -1.959    0.051    -1.679     0.002
    SECTOR1     1.024         0.504        0.078     2.032    0.043     0.034     2.014
    SECTOR2     0.614         0.957        0.025     0.642    0.521    -1.266     2.495
  ESTCIVIL1     0.531         0.420        0.049     1.264    0.207    -0.294     1.356
----------------------------------------------------------------------------------------
```

modelo3 <- lm(SALARIO ~ EDUCACION + REGION + SEXO + EXPERIENCIA + SECTOR + ESTCIVIL) == modelo1

```
> modelo |> ols_step_forward_aic()
```

Stepwise Summary

Step	Variable	AIC	SBC	SBIC	R2	Adj. R2
0	Base Model	3266.610	3275.171	1750.543	0.00000	0.00000
1	EDUCACION	3184.417	3197.258	1668.446	0.14586	0.14426
2	EXPERIENCIA	3150.099	3167.220	1634.263	0.20202	0.19902
3	SEXO	3116.735	3138.137	1601.287	0.25316	0.24893
4	REGION	3114.403	3140.086	1599.030	0.25919	0.25359
5	SECTOR	3113.929	3148.172	1596.668	0.26537	0.25701

Final Model Output

Model Summary

R	0.515	RMSE	4.401
R-Squared	0.265	MSE	19.365
Adj. R-Squared	0.257	Coef. Var	49.088
Pred R-Squared	0.247	AIC	3113.929
MAE	3.154	SBC	3148.172

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	3735.570	6	622.595	31.728	0.0000
Residual	10341.129	527	19.623		
Total	14076.699	533			

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	-4.113	1.253		-3.283	0.001	-6.574	-1.652
EDUCACION	0.940	0.080	0.478	11.687	0.000	0.782	1.098
EXPERIENCIA	0.110	0.017	0.264	6.578	0.000	0.077	0.143
SEXO1	-2.252	0.393	-0.219	-5.730	0.000	-3.025	-1.480
REGION1	-0.828	0.428	-0.073	-1.933	0.054	-1.669	0.013
SECTOR1	1.041	0.504	0.079	2.066	0.039	0.051	2.032
SECTOR2	0.650	0.957	0.026	0.679	0.497	-1.230	2.531

modelo4 <- lm(SALARIO ~ EDUCACION + REGION + SEXO + EXPERIENCIA + SECTOR) == modelo2

```
# A tibble: 9 x 5
  term        estimate std.error statistic      p.value
  <chr>          <dbl>     <dbl>     <dbl>        <dbl>
1 (Intercept)   -1.84      6.80     -0.271 0.787
2 EDUCACION      1.33      1.12      1.19  0.233
3 REGION1       -0.835     0.429    -1.95  0.0519
4 SEXO1         -2.26      0.394    -5.73  0.0000000166
5 EXPERIENCIA    0.505     1.12      0.453 0.651
6 EDAD          -0.401     1.11     -0.360 0.719
7 SECTOR1        1.03      0.504     2.03  0.0424
8 SECTOR2        0.611     0.958     0.638 0.524
9 ESTCIVIL1      0.538     0.421     1.28  0.202
```

Hay multicolinealidad

```
modelo2 |> tidy()
```

```
# A tibble: 8 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)   -4.87     1.31       -3.72 2.22e- 4
2 EDUCACION      0.829    0.0763     10.9  6.19e-25
3 REGION1       -0.840    0.428      -1.96 5.04e- 2
4 SEXO1         -2.25     0.393      -5.72 1.77e- 8
5 EDAD           0.104    0.0173      5.98 4.23e- 9
6 SECTOR1        1.02     0.504       2.03 4.27e- 2
7 SECTOR2        0.616    0.957       0.643 5.20e- 1
8 ESTCIVIL1      0.530    0.420       1.26 2.08e- 1
```