act-regresion-lineal-2

September 5, 2023

#Actividad Regresion Lineal 2 Ernesto Reynoso Lizárraga A01639915

0.0.1 Base de datos completa. No se observan valores faltantes. En caso de haberlos se realiza imputación simple.

[152]:
```
import pandas as pd
import numpy as np
import scipy.stats as stats
import statsmodels.api as sm
import matplotlib.pyplot as plt
df= pd.read_csv("/content/drive/MyDrive/Inteligencia Artificial/breast_cancer.
 ↪csv")
df.head()
```

[152]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	\
0	842302	M	17.99	10.38	122.80	1001.0	
1	842517	M	20.57	17.77	132.90	1326.0	
2	84300903	M	19.69	21.25	130.00	1203.0	
3	84348301	M	11.42	20.38	77.58	386.1	
4	84358402	M	20.29	14.34	135.10	1297.0	

	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	\
0	0.11840	0.27760	0.3001	0.14710	
1	0.08474	0.07864	0.0869	0.07017	
2	0.10960	0.15990	0.1974	0.12790	
3	0.14250	0.28390	0.2414	0.10520	
4	0.10030	0.13280	0.1980	0.10430	

	…	radius_worst	texture_worst	perimeter_worst	area_worst	\
0	…	25.38	17.33	184.60	2019.0	
1	…	24.99	23.41	158.80	1956.0	
2	…	23.57	25.53	152.50	1709.0	
3	…	14.91	26.50	98.87	567.7	
4	…	22.54	16.67	152.20	1575.0	

	smoothness_worst	compactness_worst	concavity_worst	concave points_worst	\
0	0.1622	0.6656	0.7119	0.2654	
1	0.1238	0.1866	0.2416	0.1860	

```
2            0.1444             0.4245           0.4504                0.2430
3            0.2098             0.8663           0.6869                0.2575
4            0.1374             0.2050           0.4000                0.1625

   symmetry_worst  fractal_dimension_worst
0          0.4601                  0.11890
1          0.2750                  0.08902
2          0.3613                  0.08758
3          0.6638                  0.17300
4          0.2364                  0.07678

[5 rows x 32 columns]
```

[153]:
```
df = df.drop(['id','diagnosis'],axis=1)
```

###Mostrar que las variables regresoras son independientes. En caso de no serlo realizar el procedimiento correspondiente.

[154]:
```
correlacion = df.corr()
```

[155]:
```
alta_corr=np.where((correlacion > 0.95) & (correlacion < 1))
```

[156]:
```
alta_corr
```

[156]:
```
(array([ 0,  0,  0,  0,  2,  2,  2,  2,  3,  3,  3,  3,  3, 10, 10, 12, 13,
        20, 20, 20, 20, 20, 22, 22, 22, 22, 22, 23, 23, 23]),
 array([ 2,  3, 20, 22,  0,  3, 20, 22,  0,  2, 20, 22, 23, 12, 13, 10, 10,
         0,  2,  3, 22, 23,  0,  2,  3, 20, 23,  3, 20, 22]))
```

[157]:
```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
```

[158]:
```
df_estandar = scaler.fit_transform(df)
```

[159]:
```
df_estandar = pd.DataFrame(df_estandar,columns=df.columns)
df_estandar
```

[159]:
```
     radius_mean  texture_mean  perimeter_mean  area_mean  smoothness_mean  \
0       1.097064     -2.073335        1.269934   0.984375         1.568466
1       1.829821     -0.353632        1.685955   1.908708        -0.826962
2       1.579888      0.456187        1.566503   1.558884         0.942210
3      -0.768909      0.253732       -0.592687  -0.764464         3.283553
4       1.750297     -1.151816        1.776573   1.826229         0.280372
..           …             …               …          …                …
564     2.110995      0.721473        2.060786   2.343856         1.041842
565     1.704854      2.085134        1.615931   1.723842         0.102458
566     0.702284      2.045574        0.672676   0.577953        -0.840484
```

```
567     1.838341      2.336457        1.982524   1.735218         1.525767
568    -1.808401      1.221792       -1.814389  -1.347789        -3.112085

     compactness_mean  concavity_mean  concave points_mean  symmetry_mean  \
0            3.283515        2.652874             2.532475       2.217515
1           -0.487072       -0.023846             0.548144       0.001392
2            1.052926        1.363478             2.037231       0.939685
3            3.402909        1.915897             1.451707       2.867383
4            0.539340        1.371011             1.428493      -0.009560
..                …               …                    …              …
564          0.219060        1.947285             2.320965      -0.312589
565         -0.017833        0.693043             1.263669      -0.217664
566         -0.038680        0.046588             0.105777      -0.809117
567          3.272144        3.296944             2.658866       2.137194
568         -1.150752       -1.114873            -1.261820      -0.820070

     fractal_dimension_mean  …  radius_worst  texture_worst  \
0                  2.255747  …      1.886690      -1.359293
1                 -0.868652  …      1.805927      -0.369203
2                 -0.398008  …      1.511870      -0.023974
3                  4.910919  …     -0.281464       0.133984
4                 -0.562450  …      1.298575      -1.466770
..                      …  …           …              …
564               -0.931027  …      1.901185       0.117700
565               -1.058611  …      1.536720       2.047399
566               -0.895587  …      0.561361       1.374854
567                1.043695  …      1.961239       2.237926
568               -0.561032  …     -1.410893       0.764190

     perimeter_worst  area_worst  smoothness_worst  compactness_worst  \
0           2.303601    2.001237          1.307686           2.616665
1           1.535126    1.890489         -0.375612          -0.430444
2           1.347475    1.456285          0.527407           1.082932
3          -0.249939   -0.550021          3.394275           3.893397
4           1.338539    1.220724          0.220556          -0.313395
..               …           …                 …                  …
564         1.752563    2.015301          0.378365          -0.273318
565         1.421940    1.494959         -0.691230          -0.394820
566         0.579001    0.427906         -0.809587           0.350735
567         2.303601    1.653171          1.430427           3.904848
568        -1.432735   -1.075813         -1.859019          -1.207552

     concavity_worst  concave points_worst  symmetry_worst  \
0           2.109526              2.296076        2.750622
1          -0.146749              1.087084       -0.243890
2           0.854974              1.955000        1.152255
3           1.989588              2.175786        6.046041
```

```
4           0.613179              0.729259       -0.868353
..               …                     …              …
564         0.664512              1.629151       -1.360158
565         0.236573              0.733827       -0.531855
566         0.326767              0.414069       -1.104549
567         3.197605              2.289985        1.919083
568        -1.305831             -1.745063       -0.048138

     fractal_dimension_worst
0                   1.937015
1                   0.281190
2                   0.201391
3                   4.935010
4                  -0.397100
..                       …
564                -0.709091
565                -0.973978
566                -0.318409
567                 2.219635
568                -0.751207

[569 rows x 30 columns]
```

[160]:
```
from sklearn.model_selection import train_test_split
entrenamiento, prueba = train_test_split(df_estandar, test_size=0.2,␣
 ↪random_state=42)
entrenamiento
```

[160]:
```
     radius_mean  texture_mean  perimeter_mean  area_mean  smoothness_mean  \
68     -1.447987     -0.456023       -1.366651  -1.150124         0.728714
181     1.977508      1.694187        2.089619   1.866047         1.262455
63     -1.407089     -1.263516       -1.349763  -1.120545        -1.362838
248    -0.987600      1.380033       -0.986877  -0.875668         0.014925
60     -1.123927     -1.026155       -1.129395  -0.975496         1.212639
..           …             …               …          …                …
71     -1.488033     -1.082004       -1.366651  -1.168611         0.104593
106    -0.706426     -0.223317       -0.691956  -0.689379         1.269571
270     0.046211     -0.574704       -0.068748  -0.063392        -2.282296
435    -0.041833      0.076875       -0.034972  -0.157532         0.686015
102    -0.553058      0.286311       -0.607516  -0.557982        -1.155035

     compactness_mean  concavity_mean  concave points_mean  symmetry_mean  \
68           0.700428        2.814833            -0.133333       1.093024
181          3.389643        2.007548             2.596960       2.129892
63          -0.318972       -0.363081            -0.699511       1.932741
248         -0.606466       -0.816190            -0.845247       0.311723
60          -0.449737       -0.978777            -0.929077       3.400421
```

```
..                …            …                  …            …
71           0.924055      -0.034392           -0.521016      0.329977
106         -0.050051      -0.227236           -0.362899     -0.038768
270         -1.470464      -1.023849           -1.100607     -1.108494
435          0.169787       0.298817            0.405245     -0.520693
102         -1.212155      -0.815688           -0.805266     -0.265127

     fractal_dimension_mean  …  radius_worst  texture_worst  \
68                 2.503828  …     -1.234044      -0.492965
181                1.585220  …      2.155897       1.270634
63                 0.968562  …     -1.296169      -1.049890
248                0.069801  …     -0.832304       1.549097
60                 0.964310  …     -1.087016      -1.339752
..                       …  …            …             …
71                 3.827870  …     -1.353531      -1.629614
106                0.340564  …     -0.648001       0.583433
270               -1.281175  …     -0.281464      -0.818652
435                0.374586  …      0.159621       0.834212
102               -0.854476  …     -0.606584       1.166414

     perimeter_worst  area_worst  smoothness_worst  compactness_worst  \
68         -1.243893   -0.977194          0.693984           1.159269
181         2.062335    2.124291          0.733436           3.207003
63         -1.241212   -1.002860         -1.490797          -0.550038
248        -0.872165   -0.746907          0.768505          -0.728158
60         -1.114026   -0.900022         -0.213419          -0.989865
..               …          …                …                 …
71         -1.331463   -1.048038         -0.511503          -0.067845
106        -0.647878   -0.630885          1.597003           0.074651
270        -0.381891   -0.344521         -2.047074          -1.297121
435         0.197742   -0.019835          1.268234           0.652266
102        -0.675579   -0.585004         -0.879725          -1.053734

     concavity_worst  concave points_worst  symmetry_worst  \
68          4.700669              0.919592        2.147190
181         1.946890              2.675218        1.936879
63         -0.635617             -0.970486        0.616770
248        -0.766109             -0.810759        0.822228
60         -1.201820             -1.352369        1.061659
..               …                     …             …
71         -0.617866             -1.016318       -1.046309
106         0.072498              0.109537       -0.153294
270        -1.120358             -1.237560       -0.716282
435         0.646282              1.036837        0.450138
102        -0.756514             -0.613574       -0.334485

     fractal_dimension_worst
```

```
68                  1.859432
181                 2.463465
63                  0.052877
248                -0.137199
60                 -0.207578
..                       …
71                  1.355149
106                 0.389251
270                -1.260478
435                 1.194443
102                -0.840426

[455 rows x 30 columns]
```

[161]:
```
entrenamiento.columns
```

[161]:
```
Index(['radius_mean', 'texture_mean', 'perimeter_mean', 'area_mean',
       'smoothness_mean', 'compactness_mean', 'concavity_mean',
       'concave points_mean', 'symmetry_mean', 'fractal_dimension_mean',
       'radius_se', 'texture_se', 'perimeter_se', 'area_se', 'smoothness_se',
       'compactness_se', 'concavity_se', 'concave points_se', 'symmetry_se',
       'fractal_dimension_se', 'radius_worst', 'texture_worst',
       'perimeter_worst', 'area_worst', 'smoothness_worst',
       'compactness_worst', 'concavity_worst', 'concave points_worst',
       'symmetry_worst', 'fractal_dimension_worst'],
      dtype='object')
```

0.0.2 Hipótesis nula de los coeficientes de regresión. Estadístico de prueba, distribución del estadístico de prueba.

[162]:
```
import statsmodels.formula.api as smf
modelo = smf.ols(formula =␣
 ↪'radius_mean~texture_mean+perimeter_mean+area_mean+compactness_mean+concavity_mean+symmetry
modelo = modelo.fit()
print(modelo.summary())
```

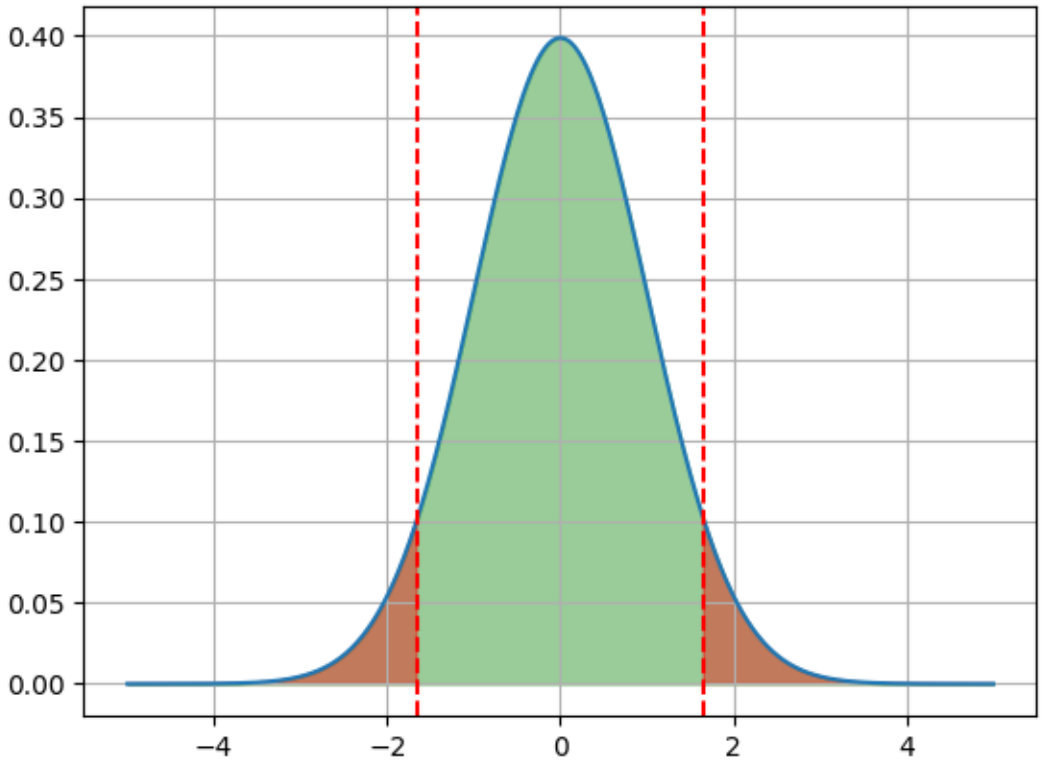
```
                            OLS Regression Results
==============================================================================
Dep. Variable:            radius_mean   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 6.725e+04
Date:                Mon, 04 Sep 2023   Prob (F-statistic):               0.00
Time:                        22:13:15   Log-Likelihood:                1235.3
No. Observations:                 455   AIC:                            -2419.
Df Residuals:                     429   BIC:                            -2311.
Df Model:                          25
Covariance Type:            nonrobust
```

```
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0004      0.001      0.549      0.583      -0.001       0.002
texture_mean  -0.0012      0.003     -0.435      0.664      -0.006       0.004
perimeter_mean 0.9534      0.018     53.796      0.000       0.919       0.988
area_mean      0.0693      0.014      5.084      0.000       0.043       0.096
compactness_mean -0.0536   0.005    -11.325      0.000      -0.063      -0.044
concavity_mean -0.0368     0.004     -8.853      0.000      -0.045      -0.029
symmetry_mean  0.0046      0.002      3.000      0.003       0.002       0.008
fractal_dimension_mean 0.0089 0.003   2.946      0.003       0.003       0.015
radius_se      0.0066      0.006      1.019      0.309      -0.006       0.019
texture_se     0.0005      0.002      0.294      0.769      -0.003       0.004
perimeter_se  -0.0186      0.006     -3.129      0.002      -0.030      -0.007
area_se        0.0015      0.004      0.336      0.737      -0.007       0.010
smoothness_se  0.0002      0.001      0.117      0.907      -0.003       0.003
compactness_se -0.0023     0.003     -0.829      0.408      -0.008       0.003
concavity_se   0.0146      0.002      6.454      0.000       0.010       0.019
symmetry_se    0.0052      0.002      2.937      0.003       0.002       0.009
fractal_dimension_se -0.0036 0.002   -1.583      0.114      -0.008       0.001
radius_worst   0.2298      0.018     12.520      0.000       0.194       0.266
texture_worst -0.0009      0.003     -0.267      0.790      -0.007       0.006
perimeter_worst -0.1123    0.015     -7.438      0.000      -0.142      -0.083
area_worst    -0.0852      0.013     -6.385      0.000      -0.111      -0.059
```

```
smoothness_worst            0.0005      0.002      0.268      0.789      -0.003
0.004
compactness_worst           0.0143      0.005      3.146      0.002       0.005
0.023
concavity_worst             0.0010      0.004      0.254      0.799      -0.007
0.009
symmetry_worst             -0.0060      0.002     -2.598      0.010      -0.010
-0.001
fractal_dimension_worst    -0.0045      0.003     -1.335      0.182      -0.011
0.002
==============================================================================
Omnibus:                       42.755   Durbin-Watson:                   2.064
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              196.966
Skew:                           0.206   Prob(JB):                     1.70e-43
Kurtosis:                       6.197   Cond. No.                         118.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

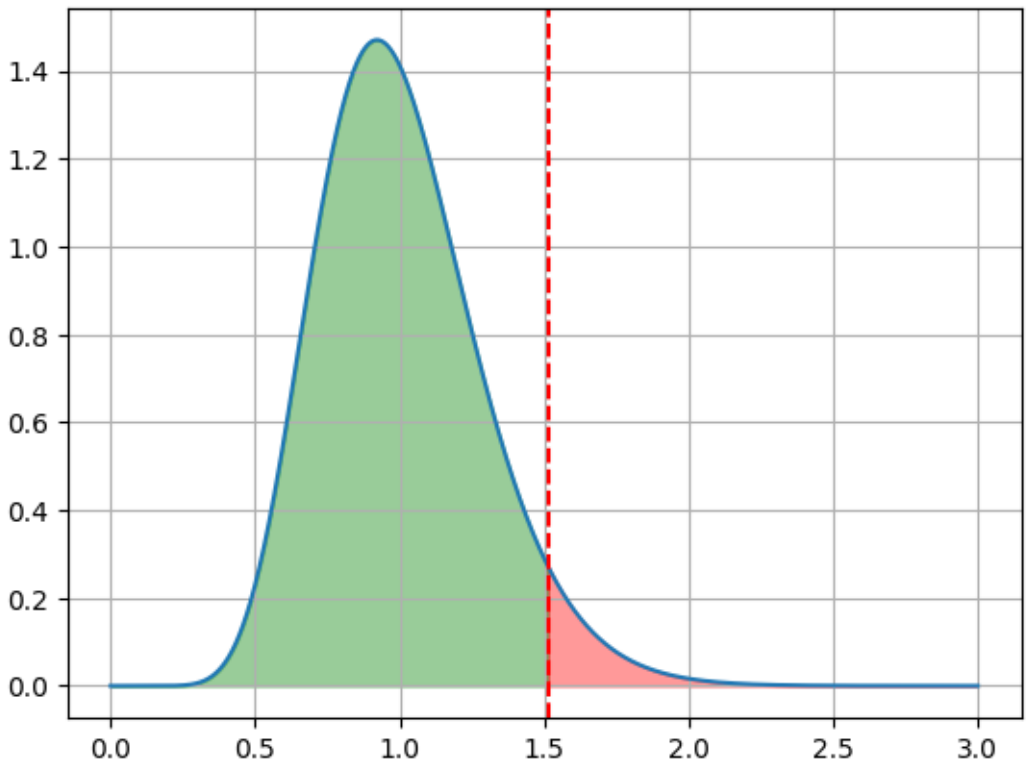
El modelo nos muestra que tiene una distribucion t-student

[185]:
```
alpha = 1 - 0.95

crit_value = stats.t.ppf(1-alpha, entrenamiento.shape[0]-1)

lim_negativo = -crit_value
lim_positivo = crit_value

x = np.linspace(-5,5,1000)

pdf = stats.t.pdf(x, entrenamiento.shape[0]-1)
plt.plot(x, pdf)

plt.fill_between(x, pdf, where=(x < crit_value) | (x > -crit_value),␣
 ↪color='green', alpha=0.4)
plt.fill_between(x, pdf, where=(x >= crit_value) | (x <= -crit_value),␣
 ↪color='red', alpha=0.4)

plt.axvline (lim_negativo, color = 'red', linestyle='--')
plt.axvline (lim_positivo, color = 'red', linestyle='--')
plt.grid(True)
plt.show()
```



0.0.3 Hipótesis nula de la significancia del modelo (prueba F-Fisher). Menciona que distribución tiene el estadístico de prueba con qué número de grados de libertad. Para un 95% de confianza realiza un diagrama en donde se muestre la distribución del estadístico de prueba, la zona de aceptación y la zona de rechazo.

[189]:
```
confianza =0.95

crit_value = stats.f.ppf(confianza, 27, entrenamiento.shape[0]-28)
x = np.linspace(0,3,10000)

pdf = stats.f.pdf(x, dfn= 27, dfd= entrenamiento.shape[0]-28)
plt.plot(x, pdf)

plt.fill_between(x, pdf, where=(x < crit_value), color='green', alpha=0.4)
plt.fill_between(x, pdf, where=(x >= crit_value), color='red', alpha=0.4)

plt.axvline (crit_value, color = 'red', linestyle='--')
#plt.axvline (modelo.fvalue, color = 'red', linestyle='--')
plt.grid(True)
plt.show()
```



0.0.4 Realiza un modelo de regresión hacia atrás (backward). Explica el criterio para ir eliminando variables del modelo.

[165]:
```
modelo = smf.ols(formula =␣
 ↪'radius_mean~texture_mean+perimeter_mean+area_mean+compactness_mean+concavity_mean+symmetry
modelo = modelo.fit()
print(modelo.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            radius_mean   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 7.022e+04
Date:                Mon, 04 Sep 2023   Prob (F-statistic):               0.00
Time:                        22:13:15   Log-Likelihood:                1235.3
No. Observations:                 455   AIC:                            -2421.
Df Residuals:                     430   BIC:                            -2318.
Df Model:                          24
Covariance Type:            nonrobust
==============================================================================
==========
                               coef    std err          t      P>|t|      [0.025
```

```
0.975]
------------------------------------------------------------------------------
-----------
Intercept                 0.0004      0.001      0.551      0.582      -0.001
0.002
texture_mean             -0.0011      0.003     -0.421      0.674      -0.006
0.004
perimeter_mean            0.9532      0.018     54.033      0.000       0.919
0.988
area_mean                 0.0695      0.014      5.134      0.000       0.043
0.096
compactness_mean         -0.0536      0.005    -11.372      0.000      -0.063
-0.044
concavity_mean           -0.0367      0.004     -8.894      0.000      -0.045
-0.029
symmetry_mean             0.0047      0.002      3.012      0.003       0.002
0.008
fractal_dimension_mean    0.0089      0.003      2.947      0.003       0.003
0.015
radius_se                 0.0066      0.006      1.027      0.305      -0.006
0.019
texture_se                0.0005      0.002      0.308      0.758      -0.003
0.004
perimeter_se             -0.0187      0.006     -3.144      0.002      -0.030
-0.007
area_se                   0.0015      0.004      0.333      0.739      -0.007
0.010
compactness_se           -0.0022      0.003     -0.835      0.404      -0.007
0.003
concavity_se              0.0145      0.002      6.468      0.000       0.010
0.019
symmetry_se               0.0052      0.002      3.015      0.003       0.002
0.009
fractal_dimension_se     -0.0036      0.002     -1.582      0.114      -0.008
0.001
radius_worst              0.2297      0.018     12.543      0.000       0.194
0.266
texture_worst            -0.0010      0.003     -0.291      0.772      -0.007
0.006
perimeter_worst          -0.1122      0.015     -7.457      0.000      -0.142
-0.083
area_worst               -0.0852      0.013     -6.397      0.000      -0.111
-0.059
smoothness_worst          0.0006      0.001      0.457      0.648      -0.002
0.003
compactness_worst         0.0142      0.005      3.155      0.002       0.005
0.023
concavity_worst           0.0010      0.004      0.245      0.807      -0.007
```

```
0.009
symmetry_worst             -0.0060      0.002     -2.662      0.008      -0.010
-0.002
fractal_dimension_worst    -0.0045      0.003     -1.338      0.182      -0.011
0.002
==============================================================================
Omnibus:                       42.874   Durbin-Watson:                   2.064
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              198.281
Skew:                           0.206   Prob(JB):                     8.79e-44
Kurtosis:                       6.208   Cond. No.                         118.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[167]:
```
modelo = smf.ols(formula =␣
 ↪'radius_mean~texture_mean+perimeter_mean+area_mean+compactness_mean+concavity_mean+symmetry
modelo = modelo.fit()
print(modelo.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            radius_mean   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 7.343e+04
Date:                Mon, 04 Sep 2023   Prob (F-statistic):               0.00
Time:                        22:16:49   Log-Likelihood:                1235.2
No. Observations:                 455   AIC:                            -2422.
Df Residuals:                     431   BIC:                            -2324.
Df Model:                          23
Covariance Type:            nonrobust
==============================================================================
==========
                             coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
----------
Intercept                    0.0004      0.001      0.560      0.576      -0.001
0.002
texture_mean                -0.0011      0.003     -0.431      0.667      -0.006
0.004
perimeter_mean               0.9533      0.018     54.114      0.000       0.919
0.988
area_mean                    0.0692      0.013      5.135      0.000       0.043
0.096
compactness_mean            -0.0539      0.005    -11.834      0.000      -0.063
-0.045
```

```
concavity_mean             -0.0361      0.003    -10.914      0.000      -0.043
-0.030
symmetry_mean               0.0047      0.002      3.015      0.003       0.002
0.008
fractal_dimension_mean      0.0088      0.003      2.940      0.003       0.003
0.015
radius_se                   0.0065      0.006      1.016      0.310      -0.006
0.019
texture_se                  0.0005      0.002      0.296      0.767      -0.003
0.004
perimeter_se               -0.0186      0.006     -3.142      0.002      -0.030
-0.007
area_se                     0.0015      0.004      0.330      0.742      -0.007
0.010
compactness_se             -0.0023      0.003     -0.872      0.384      -0.008
0.003
concavity_se                0.0148      0.002      7.453      0.000       0.011
0.019
symmetry_se                 0.0053      0.002      3.028      0.003       0.002
0.009
fractal_dimension_se       -0.0037      0.002     -1.680      0.094      -0.008
0.001
radius_worst                0.2297      0.018     12.563      0.000       0.194
0.266
texture_worst              -0.0009      0.003     -0.282      0.778      -0.007
0.006
perimeter_worst            -0.1121      0.015     -7.463      0.000      -0.142
-0.083
area_worst                 -0.0851      0.013     -6.401      0.000      -0.111
-0.059
smoothness_worst            0.0006      0.001      0.471      0.638      -0.002
0.003
compactness_worst           0.0148      0.004      3.686      0.000       0.007
0.023
symmetry_worst             -0.0060      0.002     -2.680      0.008      -0.010
-0.002
fractal_dimension_worst    -0.0044      0.003     -1.318      0.188      -0.011
0.002
==============================================================================
Omnibus:                       43.160   Durbin-Watson:                   2.063
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              199.550
Skew:                           0.211   Prob(JB):                     4.66e-44
Kurtosis:                       6.217   Cond. No.                         114.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[168]:
```
modelo = smf.ols(formula =␣
  ↪'radius_mean~texture_mean+perimeter_mean+area_mean+compactness_mean+concavity_mean+symmetry
modelo = modelo.fit()
print(modelo.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            radius_mean   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 7.693e+04
Date:                Mon, 04 Sep 2023   Prob (F-statistic):               0.00
Time:                        22:18:13   Log-Likelihood:                1235.2
No. Observations:                 455   AIC:                            -2424.
Df Residuals:                     432   BIC:                            -2330.
Df Model:                          22
Covariance Type:            nonrobust
==============================================================================
===========
                               coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
-----------
Intercept                    0.0004      0.001      0.556      0.578      -0.001
0.002
texture_mean                -0.0018      0.001     -1.844      0.066      -0.004
0.000
perimeter_mean               0.9544      0.017     55.529      0.000       0.921
0.988
area_mean                    0.0687      0.013      5.152      0.000       0.042
0.095
compactness_mean            -0.0540      0.005    -11.858      0.000      -0.063
-0.045
concavity_mean              -0.0362      0.003    -10.970      0.000      -0.043
-0.030
symmetry_mean                0.0047      0.002      3.149      0.002       0.002
0.008
fractal_dimension_mean       0.0089      0.003      2.983      0.003       0.003
0.015
radius_se                    0.0069      0.006      1.094      0.274      -0.005
0.019
texture_se                   0.0002      0.001      0.139      0.890      -0.002
0.002
perimeter_se                -0.0187      0.006     -3.150      0.002      -0.030
-0.007
area_se                      0.0012      0.004      0.278      0.781      -0.007
0.010
compactness_se              -0.0023      0.003     -0.873      0.383      -0.008
```

```
0.003
concavity_se                0.0148      0.002      7.474      0.000       0.011
0.019
symmetry_se                 0.0054      0.002      3.410      0.001       0.002
0.009
fractal_dimension_se       -0.0037      0.002     -1.693      0.091      -0.008
0.001
radius_worst                0.2287      0.018     12.770      0.000       0.194
0.264
perimeter_worst            -0.1124      0.015     -7.495      0.000      -0.142
-0.083
area_worst                 -0.0844      0.013     -6.474      0.000      -0.110
-0.059
smoothness_worst            0.0005      0.001      0.426      0.671      -0.002
0.003
compactness_worst           0.0147      0.004      3.684      0.000       0.007
0.023
symmetry_worst             -0.0063      0.002     -2.994      0.003      -0.010
-0.002
fractal_dimension_worst    -0.0043      0.003     -1.315      0.189      -0.011
0.002
==============================================================================
Omnibus:                       43.160   Durbin-Watson:                   2.063
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              199.207
Skew:                           0.213   Prob(JB):                     5.53e-44
Kurtosis:                       6.214   Cond. No.                         110.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[169]:
```
modelo = smf.ols(formula =␣
  ↪'radius_mean~texture_mean+perimeter_mean+area_mean+compactness_mean+concavity_mean+symmetry
modelo = modelo.fit()
print(modelo.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            radius_mean   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 8.078e+04
Date:                Mon, 04 Sep 2023   Prob (F-statistic):               0.00
Time:                        22:19:00   Log-Likelihood:                1235.2
No. Observations:                 455   AIC:                            -2426.
Df Residuals:                     433   BIC:                            -2336.
Df Model:                          21
Covariance Type:            nonrobust
```

```
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0004      0.001      0.555      0.579      -0.001       0.002
texture_mean  -0.0017      0.001     -2.025      0.044      -0.003   -5.09e-05
perimeter_mean 0.9545      0.017     55.740      0.000       0.921       0.988
area_mean      0.0686      0.013      5.162      0.000       0.042       0.095
compactness_mean -0.0540   0.005    -12.002      0.000      -0.063      -0.045
concavity_mean -0.0361     0.003    -11.030      0.000      -0.043      -0.030
symmetry_mean  0.0048      0.001      3.241      0.001       0.002       0.008
fractal_dimension_mean 0.0089 0.003   2.997      0.003       0.003       0.015
radius_se      0.0069      0.006      1.101      0.271      -0.005       0.019
perimeter_se  -0.0186      0.006     -3.153      0.002      -0.030      -0.007
area_se        0.0011      0.004      0.260      0.795      -0.007       0.010
compactness_se -0.0023     0.003     -0.872      0.384      -0.007       0.003
concavity_se   0.0148      0.002      7.493      0.000       0.011       0.019
symmetry_se    0.0055      0.002      3.619      0.000       0.003       0.008
fractal_dimension_se -0.0037 0.002   -1.691      0.092      -0.008       0.001
radius_worst   0.2286      0.018     12.790      0.000       0.193       0.264
perimeter_worst -0.1125    0.015     -7.519      0.000      -0.142      -0.083
area_worst     -0.0842     0.013     -6.498      0.000      -0.110      -0.059
smoothness_worst 0.0006    0.001      0.442      0.659      -0.002       0.003
compactness_worst 0.0147   0.004      3.692      0.000       0.007       0.023
symmetry_worst -0.0063     0.002     -3.116      0.002      -0.010      -0.002
```

```
fractal_dimension_worst    -0.0043      0.003     -1.319      0.188      -0.011
0.002
==============================================================================
Omnibus:                       43.207   Durbin-Watson:                   2.063
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              199.450
Skew:                           0.213   Prob(JB):                     4.90e-44
Kurtosis:                       6.215   Cond. No.                         110.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[170]:
```
modelo = smf.ols(formula =␣
 ↪'radius_mean~texture_mean+perimeter_mean+area_mean+compactness_mean+concavity_mean+symmetry
modelo = modelo.fit()
print(modelo.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            radius_mean   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 8.500e+04
Date:                Mon, 04 Sep 2023   Prob (F-statistic):               0.00
Time:                        22:19:33   Log-Likelihood:                1235.2
No. Observations:                 455   AIC:                            -2428.
Df Residuals:                     434   BIC:                            -2342.
Df Model:                          20
Covariance Type:            nonrobust
==============================================================================
===========
                               coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
-----------
Intercept                    0.0004      0.001      0.558      0.577      -0.001
0.002
texture_mean                -0.0018      0.001     -2.057      0.040      -0.003
-7.81e-05
perimeter_mean               0.9560      0.016     59.034      0.000       0.924
0.988
area_mean                    0.0680      0.013      5.191      0.000       0.042
0.094
compactness_mean            -0.0542      0.004    -12.104      0.000      -0.063
-0.045
concavity_mean              -0.0361      0.003    -11.041      0.000      -0.043
-0.030
symmetry_mean                0.0047      0.001      3.251      0.001       0.002
```

```
0.008
fractal_dimension_mean      0.0089      0.003      2.990      0.003       0.003
0.015
radius_se                   0.0076      0.006      1.346      0.179      -0.004
0.019
perimeter_se               -0.0182      0.006     -3.191      0.002      -0.029
-0.007
compactness_se             -0.0023      0.003     -0.879      0.380      -0.008
0.003
concavity_se                0.0148      0.002      7.499      0.000       0.011
0.019
symmetry_se                 0.0054      0.001      3.709      0.000       0.003
0.008
fractal_dimension_se       -0.0038      0.002     -1.716      0.087      -0.008
0.001
radius_worst                0.2268      0.016     13.763      0.000       0.194
0.259
perimeter_worst            -0.1133      0.015     -7.777      0.000      -0.142
-0.085
area_worst                 -0.0824      0.011     -7.587      0.000      -0.104
-0.061
smoothness_worst            0.0006      0.001      0.467      0.641      -0.002
0.003
compactness_worst           0.0148      0.004      3.731      0.000       0.007
0.023
symmetry_worst             -0.0062      0.002     -3.170      0.002      -0.010
-0.002
fractal_dimension_worst    -0.0043      0.003     -1.302      0.194      -0.011
0.002
==============================================================================
Omnibus:                       42.820   Durbin-Watson:                   2.061
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              199.545
Skew:                           0.199   Prob(JB):                     4.67e-44
Kurtosis:                       6.220   Cond. No.                         97.3
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[171]:
```
modelo = smf.ols(formula =␣
  ↪'radius_mean~texture_mean+perimeter_mean+area_mean+compactness_mean+concavity_mean+symmetry
modelo = modelo.fit()
print(modelo.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            radius_mean   R-squared:                       1.000
```

Model:	OLS	Adj. R-squared:	1.000
Method:	Least Squares	F-statistic:	8.964e+04
Date:	Mon, 04 Sep 2023	Prob (F-statistic):	0.00
Time:	22:20:11	Log-Likelihood:	1235.0
No. Observations:	455	AIC:	-2430.
Df Residuals:	435	BIC:	-2348.
Df Model:	19		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.0004	0.001	0.544	0.587	-0.001	0.002
texture_mean	-0.0018	0.001	-2.054	0.041	-0.003	-7.57e-05
perimeter_mean	0.9544	0.016	60.312	0.000	0.923	0.986
area_mean	0.0691	0.013	5.362	0.000	0.044	0.094
compactness_mean	-0.0537	0.004	-12.362	0.000	-0.062	-0.045
concavity_mean	-0.0360	0.003	-11.067	0.000	-0.042	-0.030
symmetry_mean	0.0047	0.001	3.238	0.001	0.002	0.008
fractal_dimension_mean	0.0089	0.003	3.003	0.003	0.003	0.015
radius_se	0.0080	0.006	1.427	0.154	-0.003	0.019
perimeter_se	-0.0186	0.006	-3.315	0.001	-0.030	-0.008
compactness_se	-0.0024	0.003	-0.913	0.362	-0.008	0.003
concavity_se	0.0147	0.002	7.491	0.000	0.011	0.019
symmetry_se	0.0053	0.001	3.683	0.000	0.002	0.008
fractal_dimension_se	-0.0039	0.002	-1.800	0.073	-0.008	0.000
radius_worst	0.2275	0.016	13.872	0.000	0.195	0.260
perimeter_worst	-0.1130	0.015	-7.771	0.000	-0.142	-0.084
area_worst	-0.0831	0.011	-7.716	0.000	-0.104	-0.062

```
compactness_worst           0.0146      0.004      3.707      0.000       0.007
0.022
symmetry_worst             -0.0061      0.002     -3.145      0.002      -0.010
-0.002
fractal_dimension_worst    -0.0040      0.003     -1.246      0.213      -0.010
0.002
==============================================================================
Omnibus:                       42.420   Durbin-Watson:                   2.062
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              197.890
Skew:                           0.190   Prob(JB):                     1.07e-43
Kurtosis:                       6.208   Cond. No.                         94.8
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[172]:
```
modelo = smf.ols(formula =␣
  ↪'radius_mean~texture_mean+perimeter_mean+area_mean+compactness_mean+concavity_mean+symmetry
modelo = modelo.fit()
print(modelo.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            radius_mean   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 9.465e+04
Date:                Mon, 04 Sep 2023   Prob (F-statistic):               0.00
Time:                        22:20:42   Log-Likelihood:                1234.6
No. Observations:                 455   AIC:                            -2431.
Df Residuals:                     436   BIC:                            -2353.
Df Model:                          18
Covariance Type:            nonrobust
==============================================================================
===========
                               coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
-----------
Intercept                    0.0004      0.001      0.532      0.595      -0.001
0.002
texture_mean                -0.0018      0.001     -2.093      0.037      -0.003
-0.000
perimeter_mean               0.9541      0.016     60.317      0.000       0.923
0.985
area_mean                    0.0693      0.013      5.377      0.000       0.044
0.095
compactness_mean            -0.0545      0.004    -12.886      0.000      -0.063
```

```
-0.046
concavity_mean             -0.0352      0.003    -11.190      0.000      -0.041
-0.029
symmetry_mean               0.0046      0.001      3.199      0.001       0.002
0.007
fractal_dimension_mean      0.0093      0.003      3.159      0.002       0.003
0.015
radius_se                   0.0088      0.006      1.577      0.115      -0.002
0.020
perimeter_se               -0.0195      0.006     -3.511      0.000      -0.030
-0.009
concavity_se                0.0140      0.002      7.775      0.000       0.010
0.018
symmetry_se                 0.0050      0.001      3.569      0.000       0.002
0.008
fractal_dimension_se       -0.0051      0.002     -2.881      0.004      -0.009
-0.002
radius_worst                0.2275      0.016     13.876      0.000       0.195
0.260
perimeter_worst            -0.1116      0.014     -7.719      0.000      -0.140
-0.083
area_worst                 -0.0840      0.011     -7.839      0.000      -0.105
-0.063
compactness_worst           0.0128      0.003      3.760      0.000       0.006
0.020
symmetry_worst             -0.0058      0.002     -3.032      0.003      -0.010
-0.002
fractal_dimension_worst    -0.0032      0.003     -1.037      0.300      -0.009
0.003
==============================================================================
Omnibus:                       42.138   Durbin-Watson:                   2.059
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              193.729
Skew:                           0.194   Prob(JB):                     8.56e-43
Kurtosis:                       6.173   Cond. No.                         92.5
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[173]:
```
modelo = smf.ols(formula =␣
  ↪'radius_mean~texture_mean+perimeter_mean+area_mean+compactness_mean+concavity_mean+symmetry
modelo = modelo.fit()
print(modelo.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            radius_mean   R-squared:                       1.000
```

Model:	OLS	Adj. R-squared:	1.000
Method:	Least Squares	F-statistic:	1.002e+05
Date:	Mon, 04 Sep 2023	Prob (F-statistic):	0.00
Time:	22:21:17	Log-Likelihood:	1234.0
No. Observations:	455	AIC:	-2432.
Df Residuals:	437	BIC:	-2358.
Df Model:	17		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.0004	0.001	0.524	0.600	-0.001	0.002
texture_mean	-0.0018	0.001	-2.101	0.036	-0.003	-0.000
perimeter_mean	0.9552	0.016	60.500	0.000	0.924	0.986
area_mean	0.0684	0.013	5.321	0.000	0.043	0.094
compactness_mean	-0.0529	0.004	-13.498	0.000	-0.061	-0.045
concavity_mean	-0.0359	0.003	-11.627	0.000	-0.042	-0.030
symmetry_mean	0.0047	0.001	3.270	0.001	0.002	0.008
fractal_dimension_mean	0.0073	0.002	3.228	0.001	0.003	0.012
radius_se	0.0093	0.006	1.675	0.095	-0.002	0.020
perimeter_se	-0.0200	0.006	-3.607	0.000	-0.031	-0.009
concavity_se	0.0146	0.002	8.532	0.000	0.011	0.018
symmetry_se	0.0054	0.001	4.145	0.000	0.003	0.008
fractal_dimension_se	-0.0059	0.002	-3.771	0.000	-0.009	-0.003
radius_worst	0.2244	0.016	13.919	0.000	0.193	0.256
perimeter_worst	-0.1100	0.014	-7.652	0.000	-0.138	-0.082
area_worst	-0.0827	0.011	-7.771	0.000	-0.104	-0.062
compactness_worst	0.0104	0.002	4.212	0.000	0.006	0.015

```
symmetry_worst            -0.0062      0.002     -3.353      0.001      -0.010
-0.003
==============================================================================
Omnibus:                       43.127   Durbin-Watson:                   2.046
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              209.319
Skew:                           0.176   Prob(JB):                     3.52e-46
Kurtosis:                       6.304   Cond. No.                         91.1
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[174]:
```
modelo = smf.ols(formula =␣
 ↪'radius_mean~texture_mean+perimeter_mean+area_mean+compactness_mean+concavity_mean+symmetry
modelo = modelo.fit()
print(modelo.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            radius_mean   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 1.060e+05
Date:                Mon, 04 Sep 2023   Prob (F-statistic):               0.00
Time:                        22:21:44   Log-Likelihood:                 1232.6
No. Observations:                 455   AIC:                            -2431.
Df Residuals:                     438   BIC:                            -2361.
Df Model:                          16
Covariance Type:            nonrobust
==============================================================================
==========
                             coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
----------
Intercept                  0.0004      0.001      0.487      0.626      -0.001
0.002
texture_mean              -0.0019      0.001     -2.179      0.030      -0.004
-0.000
perimeter_mean             0.9515      0.016     60.735      0.000       0.921
0.982
area_mean                  0.0697      0.013      5.416      0.000       0.044
0.095
compactness_mean          -0.0525      0.004    -13.401      0.000      -0.060
-0.045
concavity_mean            -0.0346      0.003    -11.542      0.000      -0.041
-0.029
symmetry_mean              0.0051      0.001      3.529      0.000       0.002
```

```
0.008
fractal_dimension_mean     0.0068      0.002      3.001      0.003       0.002
0.011
perimeter_se              -0.0112      0.002     -6.232      0.000      -0.015
-0.008
concavity_se               0.0141      0.002      8.353      0.000       0.011
0.017
symmetry_se                0.0055      0.001      4.217      0.000       0.003
0.008
fractal_dimension_se      -0.0053      0.002     -3.482      0.001      -0.008
-0.002
radius_worst               0.2414      0.013     19.234      0.000       0.217
0.266
perimeter_worst           -0.1248      0.011    -10.954      0.000      -0.147
-0.102
area_worst                -0.0831      0.011     -7.795      0.000      -0.104
-0.062
compactness_worst          0.0102      0.002      4.118      0.000       0.005
0.015
symmetry_worst            -0.0065      0.002     -3.478      0.001      -0.010
-0.003
==============================================================================
Omnibus:                       45.123   Durbin-Watson:                   2.058
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              243.075
Skew:                           0.136   Prob(JB):                     1.65e-53
Kurtosis:                       6.570   Cond. No.                         85.5
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

El criterio para eliminar una variable del modelo fue el p-valor, donde se descartaba la variable con el p-valor mas alto (y que excediese 0.05) se volvia a hacer el modelo. esto se realizaba en repetidas ocasiones hasta que no hubiera variables con un p-valor mayor a 0.05

0.0.5 Comparación entre datos reales y predicción. Análisis de los resultados.

[197]:

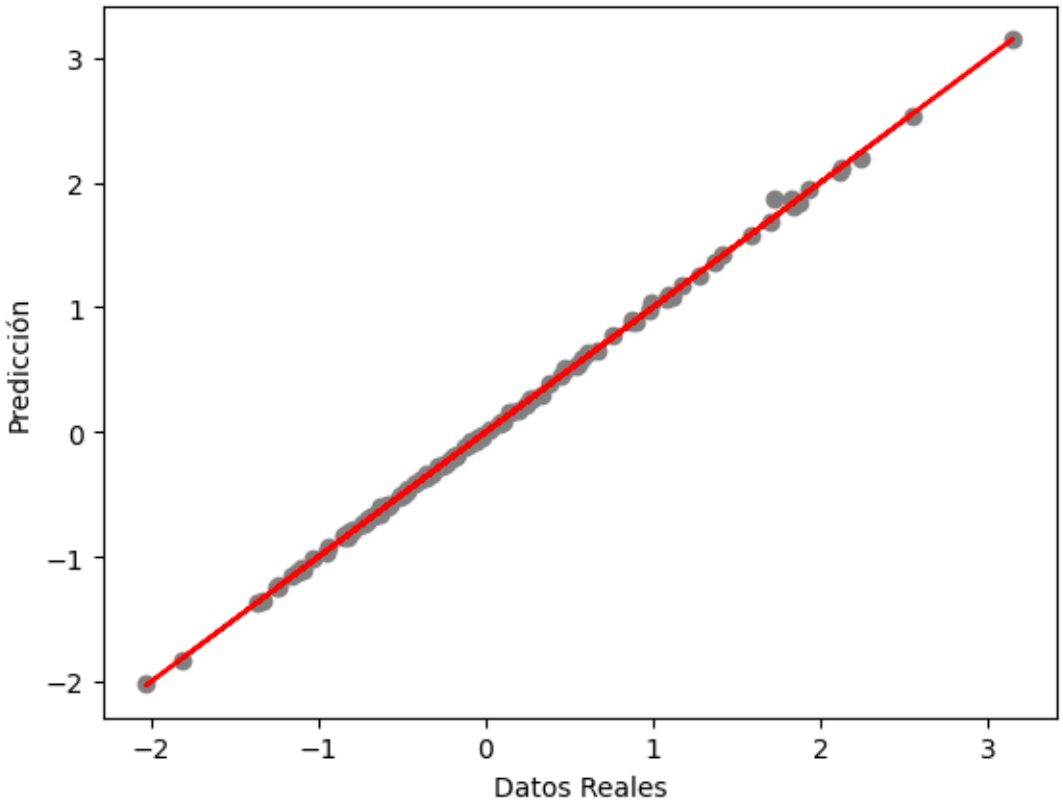
```
y_pred = modelo.params[0]+ modelo.params[1]* prueba['texture_mean'] + modelo.
 ↪params[2] * prueba['perimeter_mean'] + modelo.params[3] *␣
 ↪prueba['area_mean'] + modelo.params[4] * prueba['compactness_mean'] + modelo.
 ↪params[5] * prueba['concavity_mean'] + modelo.params[6] *␣
 ↪prueba['symmetry_mean'] + modelo.params[7] *␣
 ↪prueba['fractal_dimension_mean'] + modelo.params[8] * prueba['perimeter_se']␣
 ↪+ modelo.params[9] * prueba['concavity_se'] + modelo.params[10] *␣
 ↪prueba['symmetry_se'] + modelo.params[11] * prueba['fractal_dimension_se'] +␣
 ↪modelo.params[12] * prueba['radius_worst'] + modelo.params[13] *␣
 ↪prueba['perimeter_worst'] + modelo.params[14] * prueba['area_worst'] +␣
 ↪modelo.params[15] * prueba['compactness_worst'] + modelo.params[16] *␣
 ↪prueba['symmetry_worst']
r = prueba['radius_mean'] - y_pred
```

[198]:
```
r
```

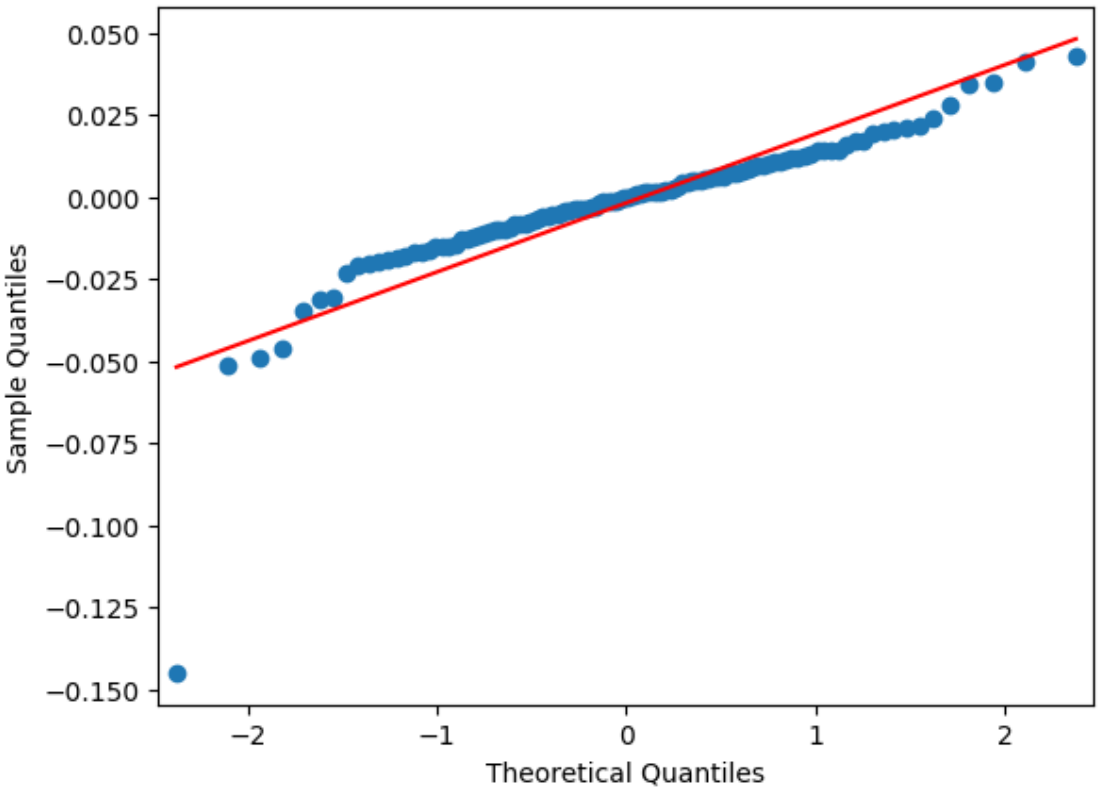
[198]:
```
204   -0.018491
70     0.001705
131   -0.015049
431    0.009664
540    0.001304
          …
486   -0.006917
75     0.002282
249   -0.001511
238    0.011553
265    0.043055
Length: 114, dtype: float64
```

[194]:
```
plt.scatter(prueba['radius_mean'],y_pred, color='gray')
plt.plot(prueba['radius_mean'], prueba['radius_mean'], color='red')
plt.xlabel("Datos Reales")
plt.ylabel("Predicción")
```

[194]:
```
Text(0, 0.5, 'Predicción')
```



[196]:
```
QQ=sm.qqplot(r,stats.norm, line='s')
```


0.050
0.025
0.000
−0.025
−0.050
−0.075
−0.100
−0.125
−0.150
Sample Quantiles
−2
−1
0
1
2
Theoretical Quantiles