Linear Regression 2

Isai Ambrocio - A01625101

Download the "breast_cancer" database available in canvas. This database contains information on the characteristics of various tumors.

Use a multiple linear regression model to predict tumor radius. The regressor variables of your model should be all the variables in the database.

1. Complete database. No missing values are observed. In case of missing values, simple imputation is performed.

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns
import statsmodels.api as sm
import scipy.stats as stats
from scipy.stats import f
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import mean_squared_error, r2_score

sns.set_theme()

df = pd.read_csv("/content/breast_cancer.csv")

df.head()
```

```
         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]
```

```
df.isna().sum()
```

```
id                         0
diagnosis                  0
radius_mean                0
texture_mean               0
perimeter_mean             0
area_mean                  0
smoothness_mean            0
compactness_mean           0
concavity_mean             0
concave points_mean        0
symmetry_mean              0
```

```
fractal_dimension_mean     0
radius_se                  0
texture_se                 0
perimeter_se               0
area_se                    0
smoothness_se              0
compactness_se             0
concavity_se               0
concave points_se          0
symmetry_se                0
fractal_dimension_se       0
radius_worst               0
texture_worst              0
perimeter_worst            0
area_worst                 0
smoothness_worst           0
compactness_worst          0
concavity_worst            0
concave points_worst       0
symmetry_worst             0
fractal_dimension_worst    0
dtype: int64
```

There is no missing data and therefore no imputations have to be made.

```
df.drop("id", axis=1, inplace=True)

df.drop("diagnosis", axis=1, inplace=True)
```

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

```
correlation_matrix = df.corr()
correlation_matrix
```

```
                         radius_mean  texture_mean  perimeter_mean
area_mean  \
radius_mean                 1.000000      0.323782        0.997855
0.987357
texture_mean                0.323782      1.000000        0.329533
0.321086
perimeter_mean              0.997855      0.329533        1.000000
0.986507
area_mean                   0.987357      0.321086        0.986507
1.000000
smoothness_mean             0.170581     -0.023389        0.207278
0.177028
compactness_mean            0.506124      0.236702        0.556936
0.498502
```

```
concavity_mean               0.676764      0.302418        0.716136
0.685983
concave points_mean          0.822529      0.293464        0.850977
0.823269
symmetry_mean                0.147741      0.071401        0.183027
0.151293
fractal_dimension_mean      -0.311631     -0.076437       -0.261477  -
0.283110
radius_se                    0.679090      0.275869        0.691765
0.732562
texture_se                  -0.097317      0.386358       -0.086761  -
0.066280
perimeter_se                 0.674172      0.281673        0.693135
0.726628
area_se                      0.735864      0.259845        0.744983
0.800086
smoothness_se               -0.222600      0.006614       -0.202694  -
0.166777
compactness_se               0.206000      0.191975        0.250744
0.212583
concavity_se                 0.194204      0.143293        0.228082
0.207660
concave points_se            0.376169      0.163851        0.407217
0.372320
symmetry_se                 -0.104321      0.009127       -0.081629  -
0.072497
fractal_dimension_se        -0.042641      0.054458       -0.005523  -
0.019887
radius_worst                 0.969539      0.352573        0.969476
0.962746
texture_worst                0.297008      0.912045        0.303038
0.287489
perimeter_worst              0.965137      0.358040        0.970387
0.959120
area_worst                   0.941082      0.343546        0.941550
0.959213
smoothness_worst             0.119616      0.077503        0.150549
0.123523
compactness_worst            0.413463      0.277830        0.455774
0.390410
concavity_worst              0.526911      0.301025        0.563879
0.512606
concave points_worst         0.744214      0.295316        0.771241
0.722017
symmetry_worst               0.163953      0.105008        0.189115
0.143570
fractal_dimension_worst      0.007066      0.119205        0.051019
0.003738
```

	smoothness_mean	compactness_mean	concavity_mean
radius_mean	0.170581	0.506124	0.676764
texture_mean	-0.023389	0.236702	0.302418
perimeter_mean	0.207278	0.556936	0.716136
area_mean	0.177028	0.498502	0.685983
smoothness_mean	1.000000	0.659123	0.521984
compactness_mean	0.659123	1.000000	0.883121
concavity_mean	0.521984	0.883121	1.000000
concave points_mean	0.553695	0.831135	0.921391
symmetry_mean	0.557775	0.602641	0.500667
fractal_dimension_mean	0.584792	0.565369	0.336783
radius_se	0.301467	0.497473	0.631925
texture_se	0.068406	0.046205	0.076218
perimeter_se	0.296092	0.548905	0.660391
area_se	0.246552	0.455653	0.617427
smoothness_se	0.332375	0.135299	0.098564
compactness_se	0.318943	0.738722	0.670279
concavity_se	0.248396	0.570517	0.691270
concave points_se	0.380676	0.642262	0.683260
symmetry_se	0.200774	0.229977	0.178009
fractal_dimension_se	0.283607	0.507318	0.449301
radius_worst	0.213120	0.535315	0.688236
texture_worst	0.036072	0.248133	0.299879
perimeter_worst	0.238853	0.590210	0.729565
area_worst	0.206718	0.509604	

```
0.675987
smoothness_worst                0.805324          0.565541
0.448822
compactness_worst               0.472468          0.865809
0.754968
concavity_worst                 0.434926          0.816275
0.884103
concave points_worst            0.503053          0.815573
0.861323
symmetry_worst                  0.394309          0.510223
0.409464
fractal_dimension_worst         0.499316          0.687382
0.514930

                         concave points_mean  symmetry_mean  \
radius_mean                         0.822529       0.147741
texture_mean                        0.293464       0.071401
perimeter_mean                      0.850977       0.183027
area_mean                           0.823269       0.151293
smoothness_mean                     0.553695       0.557775
compactness_mean                    0.831135       0.602641
concavity_mean                      0.921391       0.500667
concave points_mean                 1.000000       0.462497
symmetry_mean                       0.462497       1.000000
fractal_dimension_mean              0.166917       0.479921
radius_se                           0.698050       0.303379
texture_se                          0.021480       0.128053
perimeter_se                        0.710650       0.313893
area_se                             0.690299       0.223970
smoothness_se                       0.027653       0.187321
compactness_se                      0.490424       0.421659
concavity_se                        0.439167       0.342627
concave points_se                   0.615634       0.393298
symmetry_se                         0.095351       0.449137
fractal_dimension_se                0.257584       0.331786
radius_worst                        0.830318       0.185728
texture_worst                       0.292752       0.090651
perimeter_worst                     0.855923       0.219169
area_worst                          0.809630       0.177193
smoothness_worst                    0.452753       0.426675
compactness_worst                   0.667454       0.473200
concavity_worst                     0.752399       0.433721
concave points_worst                0.910155       0.430297
symmetry_worst                      0.375744       0.699826
fractal_dimension_worst             0.368661       0.438413

                         fractal_dimension_mean  ...  radius_worst  \
radius_mean                           -0.311631  ...      0.969539
texture_mean                          -0.076437  ...      0.352573
```

```
perimeter_mean                -0.261477  ...      0.969476
area_mean                     -0.283110  ...      0.962746
smoothness_mean                0.584792  ...      0.213120
compactness_mean               0.565369  ...      0.535315
concavity_mean                 0.336783  ...      0.688236
concave points_mean            0.166917  ...      0.830318
symmetry_mean                  0.479921  ...      0.185728
fractal_dimension_mean         1.000000  ...     -0.253691
radius_se                      0.000111  ...      0.715065
texture_se                     0.164174  ...     -0.111690
perimeter_se                   0.039830  ...      0.697201
area_se                       -0.090170  ...      0.757373
smoothness_se                  0.401964  ...     -0.230691
compactness_se                 0.559837  ...      0.204607
concavity_se                   0.446630  ...      0.186904
concave points_se              0.341198  ...      0.358127
symmetry_se                    0.345007  ...     -0.128121
fractal_dimension_se           0.688132  ...     -0.037488
radius_worst                  -0.253691  ...      1.000000
texture_worst                 -0.051269  ...      0.359921
perimeter_worst               -0.205151  ...      0.993708
area_worst                    -0.231854  ...      0.984015
smoothness_worst               0.504942  ...      0.216574
compactness_worst              0.458798  ...      0.475820
concavity_worst                0.346234  ...      0.573975
concave points_worst           0.175325  ...      0.787424
symmetry_worst                 0.334019  ...      0.243529
fractal_dimension_worst        0.767297  ...      0.093492

                         texture_worst  perimeter_worst  area_worst  \
radius_mean                   0.297008         0.965137    0.941082
texture_mean                  0.912045         0.358040    0.343546
perimeter_mean                0.303038         0.970387    0.941550
area_mean                     0.287489         0.959120    0.959213
smoothness_mean               0.036072         0.238853    0.206718
compactness_mean              0.248133         0.590210    0.509604
concavity_mean                0.299879         0.729565    0.675987
concave points_mean           0.292752         0.855923    0.809630
symmetry_mean                 0.090651         0.219169    0.177193
fractal_dimension_mean       -0.051269        -0.205151   -0.231854
radius_se                     0.194799         0.719684    0.751548
texture_se                    0.409003        -0.102242   -0.083195
perimeter_se                  0.200371         0.721031    0.730713
area_se                       0.196497         0.761213    0.811408
smoothness_se                -0.074743        -0.217304   -0.182195
compactness_se                0.143003         0.260516    0.199371
concavity_se                  0.100241         0.226680    0.188353
concave points_se             0.086741         0.394999    0.342271
symmetry_se                  -0.077473        -0.103753   -0.110343
```

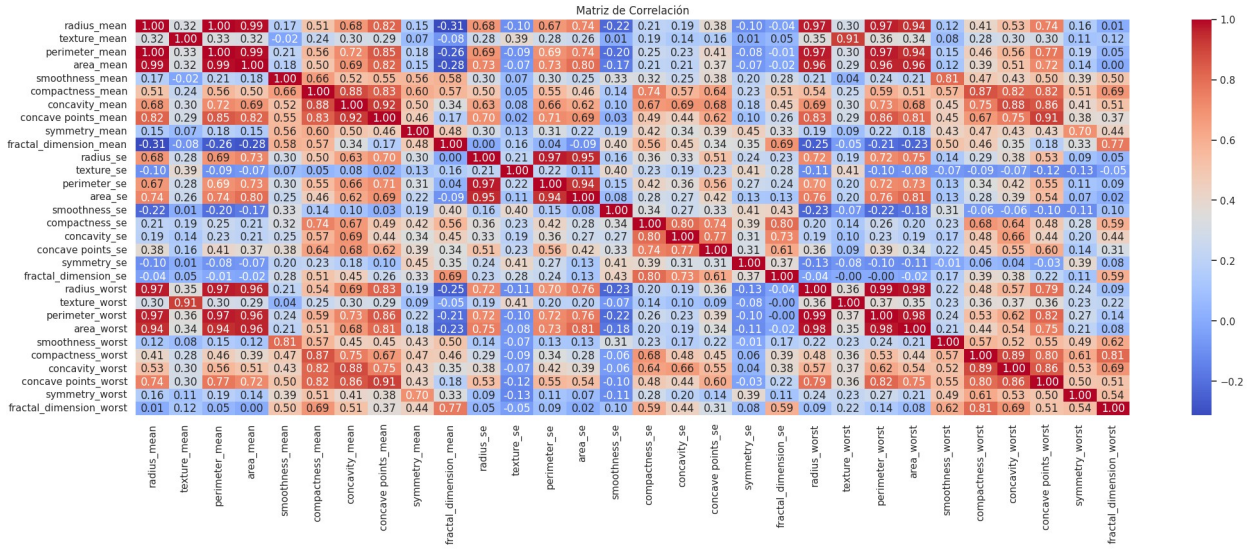
```
fractal_dimension_se          -0.003195        -0.001000   -0.022736
radius_worst                   0.359921         0.993708    0.984015
texture_worst                  1.000000         0.365098    0.345842
perimeter_worst                0.365098         1.000000    0.977578
area_worst                     0.345842         0.977578    1.000000
smoothness_worst               0.225429         0.236775    0.209145
compactness_worst              0.360832         0.529408    0.438296
concavity_worst                0.368366         0.618344    0.543331
concave points_worst           0.359755         0.816322    0.747419
symmetry_worst                 0.233027         0.269493    0.209146
fractal_dimension_worst        0.219122         0.138957    0.079647

                         smoothness_worst  compactness_worst
concavity_worst  \
radius_mean                      0.119616           0.413463
0.526911
texture_mean                     0.077503           0.277830
0.301025
perimeter_mean                   0.150549           0.455774
0.563879
area_mean                        0.123523           0.390410
0.512606
smoothness_mean                  0.805324           0.472468
0.434926
compactness_mean                 0.565541           0.865809
0.816275
concavity_mean                   0.448822           0.754968
0.884103
concave points_mean              0.452753           0.667454
0.752399
symmetry_mean                    0.426675           0.473200
0.433721
fractal_dimension_mean           0.504942           0.458798
0.346234
radius_se                        0.141919           0.287103
0.380585
texture_se                      -0.073658          -0.092439         -
0.068956
perimeter_se                     0.130054           0.341919
0.418899
area_se                          0.125389           0.283257
0.385100
smoothness_se                    0.314457          -0.055558         -
0.058298
compactness_se                   0.227394           0.678780
0.639147
concavity_se                     0.168481           0.484858
0.662564
concave points_se                0.215351           0.452888
```

```
0.549592
symmetry_se                       -0.012662           0.060255
0.037119
fractal_dimension_se               0.170568           0.390159
0.379975
radius_worst                       0.216574           0.475820
0.573975
texture_worst                      0.225429           0.360832
0.368366
perimeter_worst                    0.236775           0.529408
0.618344
area_worst                         0.209145           0.438296
0.543331
smoothness_worst                   1.000000           0.568187
0.518523
compactness_worst                  0.568187           1.000000
0.892261
concavity_worst                    0.518523           0.892261
1.000000
concave points_worst               0.547691           0.801080
0.855434
symmetry_worst                     0.493838           0.614441
0.532520
fractal_dimension_worst            0.617624           0.810455
0.686511

                          concave points_worst  symmetry_worst  \
radius_mean                           0.744214        0.163953
texture_mean                          0.295316        0.105008
perimeter_mean                        0.771241        0.189115
area_mean                             0.722017        0.143570
smoothness_mean                       0.503053        0.394309
compactness_mean                      0.815573        0.510223
concavity_mean                        0.861323        0.409464
concave points_mean                   0.910155        0.375744
symmetry_mean                         0.430297        0.699826
fractal_dimension_mean                0.175325        0.334019
radius_se                             0.531062        0.094543
texture_se                           -0.119638       -0.128215
perimeter_se                          0.554897        0.109930
area_se                               0.538166        0.074126
smoothness_se                        -0.102007       -0.107342
compactness_se                        0.483208        0.277878
concavity_se                          0.440472        0.197788
concave points_se                     0.602450        0.143116
symmetry_se                          -0.030413        0.389402
fractal_dimension_se                  0.215204        0.111094
radius_worst                          0.787424        0.243529
texture_worst                         0.359755        0.233027
```

```
perimeter_worst                  0.816322        0.269493
area_worst                       0.747419        0.209146
smoothness_worst                 0.547691        0.493838
compactness_worst                0.801080        0.614441
concavity_worst                  0.855434        0.532520
concave points_worst             1.000000        0.502528
symmetry_worst                   0.502528        1.000000
fractal_dimension_worst          0.511114        0.537848
```

```
                         fractal_dimension_worst
radius_mean                             0.007066
texture_mean                            0.119205
perimeter_mean                          0.051019
area_mean                               0.003738
smoothness_mean                         0.499316
compactness_mean                        0.687382
concavity_mean                          0.514930
concave points_mean                     0.368661
symmetry_mean                           0.438413
fractal_dimension_mean                  0.767297
radius_se                               0.049559
texture_se                             -0.045655
perimeter_se                            0.085433
area_se                                 0.017539
smoothness_se                           0.101480
compactness_se                          0.590973
concavity_se                            0.439329
concave points_se                       0.310655
symmetry_se                             0.078079
fractal_dimension_se                    0.591328
radius_worst                            0.093492
texture_worst                           0.219122
perimeter_worst                         0.138957
area_worst                              0.079647
smoothness_worst                        0.617624
compactness_worst                       0.810455
concavity_worst                         0.686511
concave points_worst                    0.511114
symmetry_worst                          0.537848
fractal_dimension_worst                 1.000000

[30 rows x 30 columns]
```

```
plt.figure(figsize=(25, 8))
sns.heatmap(correlation_matrix, annot = True, cmap = "coolwarm",
fmt=".2f")
plt.title("Matriz de Correlación")
plt.show()
```



Model

$$Y = \beta X + \epsilon$$

$$\hat{Y} = \beta^{\wedge} X$$

$$\beta X = \beta_0(1) + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

We define the dependent and independent variables of our model.

```
x = df[["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"]]


y = df["radius_mean"]
```

We add an intercept to the model. The main reason is to ensure that the regression model considers an independent or constant term in the regression equation. Likewise, we create and adjust the linear regression model.

```
high_correlation = np.where((correlation_matrix > 
0.95)&(correlation_matrix < 1))

high_correlation
```

```
(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]))
```

```
low_correlation = np.where((correlation_matrix < -
0.95)&(correlation_matrix > -1))

low_correlation
```

```
(array([], dtype=int64), array([], dtype=int64))
```

There is only positive correlation

3. Null hypothesis of regression coefficients. Test statistic, distribution of the test statistic.

For 95% confidence, make a diagram showing the distribution of the test statistic, the acceptance zone and the rejection zone.

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 
0.20)

scaler = StandardScaler()

x_train_std = scaler.fit_transform(x_train)
x_train_std
x_test_std = scaler.transform(x_test)

x_train_std = pd.DataFrame(x_train_std, columns = x.columns)

x_train_std
```

```
     texture_mean  perimeter_mean  area_mean  smoothness_mean  \
0        1.262201        1.274578   1.225251        -0.428486   
1        0.330120        1.224885   1.157012         0.716595   
2        3.347182        0.516761   0.283838         0.681916   
3       -0.095244       -0.740470  -0.671222        -0.085170   
4       -0.246329       -1.248994  -1.020094        -0.926468   
..            ...             ...        ...              ...   
450     -0.685640       -0.111855  -0.146067        -0.626847   
451     -1.329496       -0.554536  -0.568295         0.938537   
452     -0.051080       -0.650609  -0.617200        -0.538070   
453     -1.162140       -0.804657  -0.706479        -0.185738   
```

```
454     -0.460174       -0.060919  -0.102280        -1.110958

     compactness_mean  concavity_mean  concave points_mean
symmetry_mean  \
0            0.323275        0.734572             0.918587
0.515450
1            0.403390        0.860571             1.230720
0.179546
2            1.149226        1.087617             0.952445
1.013889
3           -0.933201       -0.718024            -0.666679
0.396258
4           -0.941785       -0.912261            -1.116654      -
0.470591
..                ...             ...                  ...
...
450         -0.929577       -0.909392            -0.716938      -
1.395231
451         -0.258706       -0.625085            -0.529129      -
0.553665
452         -0.660999       -0.763808            -0.900779      -
1.001537
453         -1.238593       -1.078217            -1.181091
0.519062
454         -0.954565       -0.762311            -0.409302      -
0.611455

     fractal_dimension_mean  radius_se  ...  radius_worst
texture_worst  \
0                -0.921932   0.537347  ...      1.752370
0.794547
1                -0.490537   0.521011  ...      1.191436
0.111220
2                 0.489780  -0.555081  ...      0.491307
3.899547
3                -0.537548  -0.260310  ...     -0.682500      -
0.494198
4                 0.160703  -0.201001  ...     -1.033603      -
0.145230
..                     ...        ...  ...           ...             .
..
450              -0.570732  -0.511043  ...     -0.210900
0.010588
451               0.431708  -0.521698  ...     -0.555771      -
1.581678
452              -0.330147  -0.625045  ...     -0.586934      -
0.085176
453              -0.656458  -0.570708  ...     -0.815462      -
1.059039
```

```
454               -0.735271  -0.880040  ...     -0.250373      -
0.028367

     perimeter_worst  area_worst  smoothness_worst  compactness_worst
\
0           1.397755    1.326533         -0.340930           0.431130

1           1.077636    1.098773          0.725731           0.007193

2           0.601945    0.296358          2.442045           1.923135

3          -0.729093   -0.630099         -0.340930          -1.045689

4          -1.049213   -0.850851          0.269217          -0.803349

..               ...         ...               ...                ...

450        -0.282422   -0.300373         -0.762327          -0.959636

451        -0.576214   -0.553537          0.853028          -0.564806

452        -0.627074   -0.559844         -0.204854          -0.407886

453        -0.861928   -0.714546         -0.297035          -1.132501

454        -0.291996   -0.317718         -1.144218          -0.818535


     concavity_worst  concave points_worst  symmetry_worst  \
0           1.421435              1.264597        1.697056
1           0.488026              0.579282        0.283887
2           1.397118              1.353718        0.985665
3          -0.845402             -0.770450        0.655605
4          -0.970122             -1.356809        0.256649
..               ...                   ...             ...
450        -1.022685             -0.887230       -0.856902
451        -0.511320             -0.604960       -0.185567
452        -0.581934             -0.678716       -0.579716
453        -1.246661             -1.570240       -0.222418
454        -0.874209             -0.525366       -0.598942

     fractal_dimension_worst
0                   0.461697
1                  -0.305292
2                   1.942929
3                  -0.750416
4                  -0.298833
..                       ...
450                -0.708971
451                -0.203565
452                -0.235321
```

```
453                -1.100270
454                -0.980243

[455 rows x 29 columns]
```

```
x_train_const = sm.add_constant(x_train_std)

model = sm.OLS(y_train.values, x_train_const).fit()
print(model.summary())
```

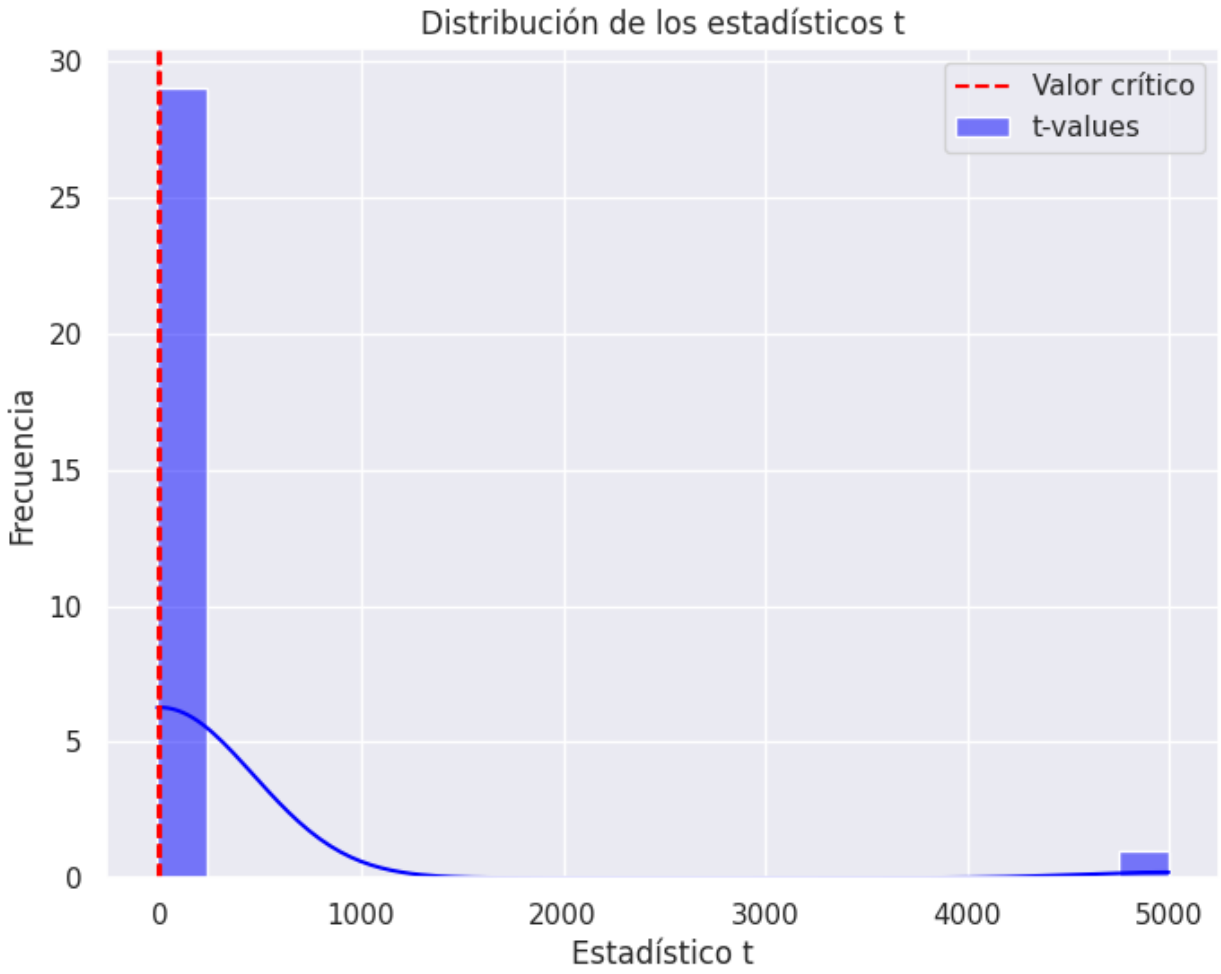
```
                            OLS Regression Results

==============================================================================
========
Dep. Variable:                          y   R-squared:
1.000
Model:                                OLS   Adj. R-squared:
1.000
Method:                     Least Squares   F-statistic:
5.381e+04
Date:                  Tue, 12 Sep 2023   Prob (F-statistic):
0.00
Time:                          05:13:27   Log-Likelihood:
651.44
No. Observations:                   455   AIC:
-1243.
Df Residuals:                       425   BIC:
-1119.
Df Model:                            29

Covariance Type:            nonrobust

==============================================================================
=====================
                               coef    std err          t      P>|t|
[0.025      0.975]
------------------------------------------------------------------------------
--------------------
const                        14.0186      0.003   4999.553      0.000
14.013      14.024
texture_mean                  0.0057      0.010      0.574      0.566
-0.014       0.025
perimeter_mean                3.2441      0.064     51.075      0.000
3.119       3.369
area_mean                     0.2950      0.049      6.015      0.000
0.199       0.391
smoothness_mean               0.0255      0.008      3.092      0.002
0.009       0.042
compactness_mean             -0.1942      0.017    -11.123      0.000
-0.228      -0.160
```

```
concavity_mean              -0.1365      0.023     -5.926      0.000
-0.182      -0.091
concave points_mean          0.0183      0.022      0.846      0.398
-0.024       0.061
symmetry_mean                0.0086      0.006      1.484      0.139
-0.003       0.020
fractal_dimension_mean       0.0133      0.012      1.156      0.248
-0.009       0.036
radius_se                    0.0249      0.025      1.014      0.311
-0.023       0.073
texture_se                   0.0061      0.006      1.014      0.311
-0.006       0.018
perimeter_se                -0.0731      0.024     -3.085      0.002
-0.120      -0.027
area_se                      0.0060      0.018      0.329      0.742
-0.030       0.042
smoothness_se                0.0072      0.006      1.266      0.206
-0.004       0.018
compactness_se              -0.0083      0.011     -0.745      0.457
-0.030       0.014
concavity_se                 0.0476      0.011      4.213      0.000
0.025       0.070
concave points_se            0.0119      0.010      1.235      0.217
-0.007       0.031
symmetry_se                  0.0006      0.007      0.084      0.933
-0.012       0.014
fractal_dimension_se        -0.0086      0.009     -0.949      0.343
-0.027       0.009
radius_worst                 0.7838      0.067     11.716      0.000
0.652       0.915
texture_worst               -0.0132      0.013     -1.039      0.299
-0.038       0.012
perimeter_worst             -0.3559      0.056     -6.334      0.000
-0.466      -0.245
area_worst                  -0.2927      0.048     -6.052      0.000
-0.388      -0.198
smoothness_worst            -0.0267      0.009     -2.886      0.004
-0.045      -0.009
compactness_worst            0.0577      0.018      3.289      0.001
0.023       0.092
concavity_worst              0.0025      0.016      0.159      0.874
-0.028       0.034
concave points_worst         0.0052      0.017      0.304      0.761
-0.028       0.038
symmetry_worst              -0.0074      0.009     -0.829      0.407
-0.025       0.010
fractal_dimension_worst     -0.0099      0.013     -0.769      0.442
-0.035       0.015
==============================================================================
```

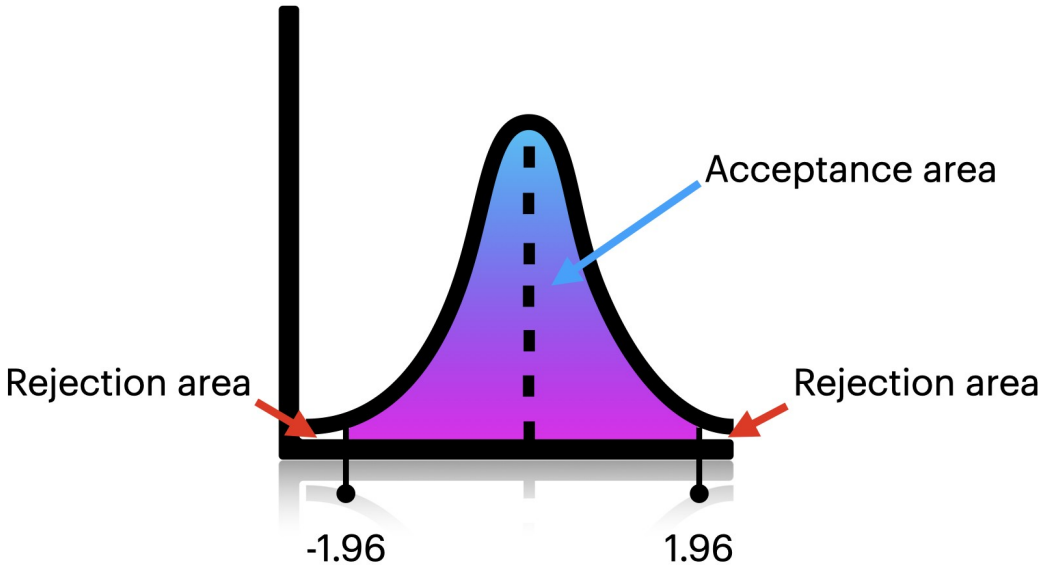
```
========
Omnibus:                       55.984   Durbin-Watson:
1.955
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
435.469
Skew:                           0.033   Prob(JB):
2.75e-95
Kurtosis:                       7.792   Cond. No.
130.
================================================================================
========

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

Calcular los grados de libertad (n - k)

```
df_degrees_freedom = len(y_train) - len(x_train_const.columns)

confidence_level = 0.95

alpha = 1 - confidence_level
t_critical = stats.t.ppf(1 - alpha / 2, df_degrees_freedom)
print("Valor crítico de t:", t_critical)

Valor crítico de t: 1.9655614588324097

plt.figure(figsize=(8, 6))
sns.histplot(model.tvalues, kde=True, bins=20, color="blue", label="t-
values")
plt.axvline(t_critical, color="red", linestyle="--", label="Valor
crítico")
plt.axvline(-t_critical, color="red", linestyle="--")
plt.xlabel("Estadístico t")
plt.ylabel("Frecuencia")
plt.title("Distribución de los estadísticos t")
plt.legend()
plt.show()
```



Confidence level of 95%.



The distribution of the test statistic is a t-distribution with n - k - 1 degrees of freedom, where n is the number of observations and k is the number of regressor variables.

Considering that we want a confidence value of 95% we know that we have 5% in the rejection area, which is divided in two due to the positive and negative area of our test statistic distribution. So each side would be 2.5% respectively.

Therefore, we observe that for the requested confidence level, our rejection areas would be less or equal to -1.96 and equal o greater than 1.96.

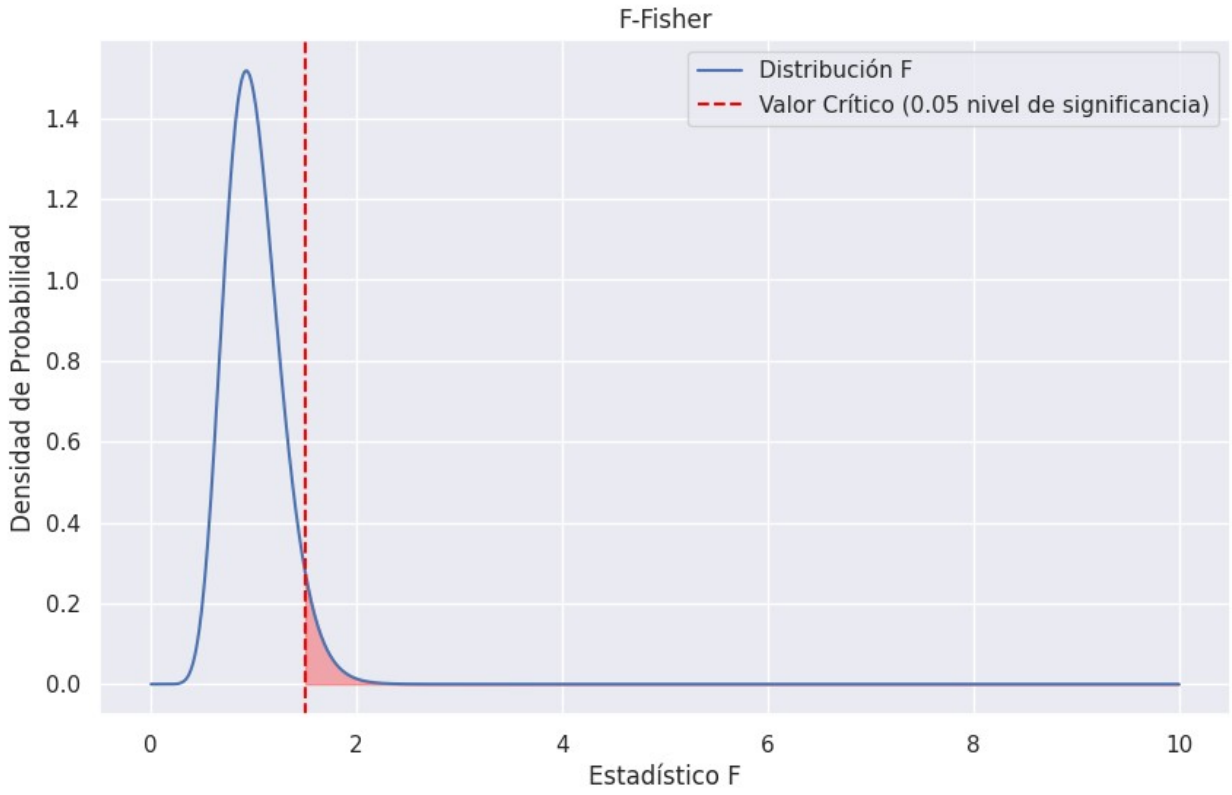
4. Hipótesis nula de la significancia del modelo (prueba F-Fisher). Menciona que distribución tiene el estadístico de prueba con qué número de grados de libertad. Para un 95% de confianza realiza un diagrama en donde se muestre la distribución del estadístico de prueba, la zona de aceptación y la zona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

The F test follows an F distribution with two degrees of freedom: one for the numerator (model degrees of freedom) and one for the denominator (error degrees of freedom). In the model summary, we can find the F statistic, the associated p-value and the degrees of freedom of the model and of the error.

The model summary provides the following relevant information:

- F-statistic (F-statistic): 5.110e+04
- Prob (F-statistic) (p-value associated with F-statistic): 0.00
- Df Model (degrees of freedom of the model): 29
- Df Residuals (error degrees of freedom): 425

```
df_model = 29
df_error = 425
alpha = 0.05

critical_value = stats.f.ppf(1 - alpha, df_model, df_error)

f_values = np.linspace(0, 10, 1000)
pdf = stats.f.pdf(f_values, df_model, df_error)

plt.figure(figsize=(10, 6))
plt.plot(f_values, pdf, label="Distribución F")
plt.axvline(critical_value, color='red', linestyle="--", label=f"Valor
Crítico ({alpha} nivel de significancia)")
plt.fill_between(f_values, 0, pdf, where=(f_values > critical_value),
color="red", alpha=0.3)
plt.xlabel("Estadístico F")
plt.ylabel("Densidad de Probabilidad")
plt.title("F-Fisher")
plt.legend()
plt.show()
```



As we can see, the calculated value exceeds the significance threshold and, therefore, we can conclude that the model is significant. In other words, we reject the null hypothesis since at least one of the regressor variables has a significant effect on the dependent variable (tumor radius) at a 95% confidence level.

5. Perform a backward regression model. Explain the criteria for eliminating variables from the model.

A backward model elimination consists of having a complete model with all the predictor variables available. Consequently, the predictor variable that has the highest P value is eliminated, this is because the variable that has the weakest relationship with the dependent variable is being eliminated and a lower one (which is retained) is more likely to be real. This leads us to the principle of parsimony, which tells us that if we can explain the same thing in a simpler way it is much better. In addition, there are two main reasons why higher values are removed, one is to improve the accuracy of the model and the other is to reduce multicollinearity.

Eliminating `texture_mean` with `P>|t|` = 0.960

```
x_backward = x_train_const.copy()
x_backward.drop(["texture_mean"], inplace=True, axis=1)

model_backward = sm.OLS(y_train.values, x_backward).fit()
print(model_backward.summary())
```

```
                            OLS Regression Results
==============================================================================
========
Dep. Variable:                          y   R-squared:
1.000
Model:                                OLS   Adj. R-squared:
1.000
Method:                     Least Squares   F-statistic:
5.582e+04
Date:                    Tue, 12 Sep 2023   Prob (F-statistic):
0.00
Time:                            05:13:28   Log-Likelihood:
651.26
No. Observations:                     455   AIC:
-1245.
Df Residuals:                         426   BIC:
-1125.
Df Model:                              28

Covariance Type:                nonrobust

==============================================================================
=====================
                                 coef    std err          t      P>|t|
[0.025      0.975]
------------------------------------------------------------------------------
---------------------
const                         14.0186      0.003   5003.494      0.000
14.013      14.024
perimeter_mean                 3.2483      0.063     51.529      0.000
3.124       3.372
area_mean                      0.2939      0.049      6.003      0.000
0.198       0.390
smoothness_mean                0.0258      0.008      3.137      0.002
0.010       0.042
compactness_mean              -0.1938      0.017    -11.118      0.000
-0.228      -0.160
concavity_mean                -0.1357      0.023     -5.907      0.000
-0.181      -0.091
concave points_mean            0.0175      0.022      0.812      0.418
-0.025       0.060
symmetry_mean                  0.0090      0.006      1.564      0.119
-0.002       0.020
fractal_dimension_mean         0.0136      0.012      1.181      0.238
-0.009       0.036
radius_se                      0.0273      0.024      1.127      0.260
-0.020       0.075
texture_se                     0.0043      0.005      0.838      0.403
-0.006       0.014
```

```
perimeter_se                -0.0731      0.024     -3.089      0.002
-0.120      -0.027
area_se                      0.0039      0.018      0.220      0.826
-0.031       0.039
smoothness_se                0.0079      0.005      1.438      0.151
-0.003       0.019
compactness_se              -0.0088      0.011     -0.791      0.429
-0.031       0.013
concavity_se                 0.0472      0.011      4.190      0.000
0.025       0.069
concave points_se            0.0125      0.010      1.317      0.189
-0.006       0.031
symmetry_se                  0.0012      0.007      0.179      0.858
-0.012       0.014
fractal_dimension_se        -0.0088      0.009     -0.964      0.336
-0.027       0.009
radius_worst                 0.7808      0.067     11.716      0.000
0.650       0.912
texture_worst               -0.0065      0.005     -1.334      0.183
-0.016       0.003
perimeter_worst             -0.3590      0.056     -6.425      0.000
-0.469      -0.249
area_worst                  -0.2896      0.048     -6.031      0.000
-0.384      -0.195
smoothness_worst            -0.0280      0.009     -3.113      0.002
-0.046      -0.010
compactness_worst            0.0584      0.017      3.343      0.001
0.024       0.093
concavity_worst              0.0024      0.016      0.153      0.879
-0.029       0.033
concave points_worst         0.0045      0.017      0.269      0.788
-0.029       0.038
symmetry_worst              -0.0084      0.009     -0.961      0.337
-0.026       0.009
fractal_dimension_worst     -0.0101      0.013     -0.790      0.430
-0.035       0.015
==============================================================================
Omnibus:                       56.382   Durbin-Watson:
1.948
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
444.883
Skew:                           0.011   Prob(JB):
2.48e-97
Kurtosis:                       7.844   Cond. No.
128.
==============================================================================

Notes:
```

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

Eliminating `concave points_mean` with `P>|t|` = 0.747

```
x_backward.drop(["concave points_mean"], inplace=True, axis=1)

model_backward = sm.OLS(y_train.values, x_backward).fit()
print(model_backward.summary())
```

```
                            OLS Regression Results

==============================================================================
========
Dep. Variable:                          y   R-squared:
1.000
Model:                                OLS   Adj. R-squared:
1.000
Method:                     Least Squares   F-statistic:
5.793e+04
Date:                   Tue, 12 Sep 2023   Prob (F-statistic):
0.00
Time:                           05:13:28   Log-Likelihood:
650.91
No. Observations:                   455   AIC:
-1246.
Df Residuals:                       427   BIC:
-1130.
Df Model:                            27

Covariance Type:            nonrobust

==============================================================================
=====================
                                coef    std err          t      P>|t|
[0.025      0.975]
------------------------------------------------------------------------------
--------------------
const                        14.0186      0.003   5005.496      0.000
14.013      14.024
perimeter_mean                3.2523      0.063     51.769      0.000
3.129       3.376
area_mean                     0.2945      0.049      6.018      0.000
0.198       0.391
smoothness_mean               0.0279      0.008      3.564      0.000
0.013       0.043
compactness_mean             -0.1935      0.017    -11.107      0.000
-0.228      -0.159
concavity_mean               -0.1225      0.016     -7.579      0.000
-0.154      -0.091
```

```
symmetry_mean                0.0091      0.006      1.586      0.113
-0.002       0.020
fractal_dimension_mean       0.0134      0.011      1.164      0.245
-0.009       0.036
radius_se                    0.0301      0.024      1.261      0.208
-0.017       0.077
texture_se                   0.0045      0.005      0.887      0.376
-0.005       0.015
perimeter_se                -0.0735      0.024     -3.105      0.002
-0.120      -0.027
area_se                      0.0009      0.018      0.052      0.959
-0.034       0.035
smoothness_se                0.0083      0.005      1.522      0.129
-0.002       0.019
compactness_se              -0.0082      0.011     -0.734      0.464
-0.030       0.014
concavity_se                 0.0438      0.010      4.192      0.000
0.023       0.064
concave points_se            0.0125      0.010      1.319      0.188
-0.006       0.031
symmetry_se                  0.0010      0.007      0.146      0.884
-0.012       0.014
fractal_dimension_se        -0.0089      0.009     -0.983      0.326
-0.027       0.009
radius_worst                 0.7803      0.067     11.714      0.000
0.649       0.911
texture_worst               -0.0067      0.005     -1.389      0.166
-0.016       0.003
perimeter_worst             -0.3630      0.056     -6.524      0.000
-0.472      -0.254
area_worst                  -0.2865      0.048     -5.988      0.000
-0.380      -0.192
smoothness_worst            -0.0291      0.009     -3.286      0.001
-0.047      -0.012
compactness_worst            0.0575      0.017      3.297      0.001
0.023       0.092
concavity_worst             -0.0016      0.015     -0.108      0.914
-0.031       0.028
concave points_worst         0.0121      0.014      0.852      0.394
-0.016       0.040
symmetry_worst              -0.0083      0.009     -0.949      0.343
-0.025       0.009
fractal_dimension_worst     -0.0100      0.013     -0.782      0.435
-0.035       0.015
==============================================================================
========
Omnibus:                       57.194   Durbin-Watson:
1.948
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
```

```
461.614
Skew:                          0.006   Prob(JB):
5.78e-101
Kurtosis:                      7.934   Cond. No.
123.
================================================================================

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

Eliminating `concave points_mean` with `P>|t|` = 0.664

```
x_backward.drop(["fractal_dimension_se"], inplace=True, axis=1)

model_backward = sm.OLS(y_train.values, x_backward).fit()
print(model_backward.summary())
```

```
                            OLS Regression Results

================================================================================
Dep. Variable:                      y   R-squared:
1.000
Model:                            OLS   Adj. R-squared:
1.000
Method:                 Least Squares   F-statistic:
6.017e+04
Date:                Tue, 12 Sep 2023   Prob (F-statistic):
0.00
Time:                        05:13:28   Log-Likelihood:
650.40
No. Observations:                 455   AIC:
-1247.
Df Residuals:                     428   BIC:
-1136.
Df Model:                          26

Covariance Type:            nonrobust

================================================================================
                                 coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
const                        14.0186      0.003   5005.688      0.000
14.013      14.024
perimeter_mean                3.2486      0.063     51.805      0.000
```

```
3.125       3.372
area_mean                0.2963      0.049      6.059       0.000
0.200       0.392
smoothness_mean          0.0284      0.008      3.640       0.000
0.013       0.044
compactness_mean        -0.1927      0.017    -11.074       0.000
-0.227      -0.159
concavity_mean          -0.1232      0.016     -7.636       0.000
-0.155      -0.092
symmetry_mean            0.0092      0.006      1.589       0.113
-0.002       0.020
fractal_dimension_mean   0.0122      0.011      1.064       0.288
-0.010       0.035
radius_se                0.0233      0.023      1.020       0.308
-0.022       0.068
texture_se               0.0043      0.005      0.848       0.397
-0.006       0.014
perimeter_se            -0.0677      0.023     -2.954       0.003
-0.113      -0.023
area_se                  0.0021      0.017      0.118       0.906
-0.032       0.036
smoothness_se            0.0084      0.005      1.541       0.124
-0.002       0.019
compactness_se          -0.0138      0.010     -1.454       0.147
-0.033       0.005
concavity_se             0.0415      0.010      4.075       0.000
0.021       0.061
concave points_se        0.0122      0.010      1.287       0.199
-0.006       0.031
symmetry_se              0.0011      0.007      0.165       0.869
-0.012       0.014
radius_worst             0.7895      0.066     11.971       0.000
0.660       0.919
texture_worst           -0.0066      0.005     -1.359       0.175
-0.016       0.003
perimeter_worst         -0.3726      0.055     -6.802       0.000
-0.480      -0.265
area_worst              -0.2865      0.048     -5.988       0.000
-0.380      -0.192
smoothness_worst        -0.0288      0.009     -3.251       0.001
-0.046      -0.011
compactness_worst        0.0631      0.016      3.823       0.000
0.031       0.095
concavity_worst          0.0013      0.015      0.087       0.931
-0.028       0.030
concave points_worst     0.0121      0.014      0.859       0.391
-0.016       0.040
symmetry_worst          -0.0080      0.009     -0.917       0.360
-0.025       0.009
```

```
fractal_dimension_worst    -0.0167      0.011     -1.548      0.122
-0.038       0.005
==============================================================================
========
Omnibus:                       57.044   Durbin-Watson:
1.945
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
456.132
Skew:                           0.040   Prob(JB):
8.96e-100
Kurtosis:                       7.904   Cond. No.
122.
==============================================================================
========

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

Eliminating `symmetry_mean` with `P>|t|` = 0.657

```
x_backward.drop(["symmetry_mean"], inplace=True, axis=1)

model_backward = sm.OLS(y_train.values, x_backward).fit()
print(model_backward.summary())
```

```
                            OLS Regression Results

==============================================================================
========
Dep. Variable:                        y   R-squared:
1.000
Model:                              OLS   Adj. R-squared:
1.000
Method:                   Least Squares   F-statistic:
6.235e+04
Date:                  Tue, 12 Sep 2023   Prob (F-statistic):
0.00
Time:                          05:13:28   Log-Likelihood:
649.06
No. Observations:                 455   AIC:
-1246.
Df Residuals:                     429   BIC:
-1139.
Df Model:                          25

Covariance Type:             nonrobust

==============================================================================
====================
```

```
                              coef    std err          t      P>|t|
[0.025      0.975]
------------------------------------------------------------------------------
---------------------
const                       14.0186      0.003   4996.808      0.000
14.013      14.024
perimeter_mean               3.2595      0.062     52.200      0.000
3.137       3.382
area_mean                    0.2899      0.049      5.938      0.000
0.194       0.386
smoothness_mean              0.0306      0.008      3.971      0.000
0.015       0.046
compactness_mean            -0.1883      0.017    -10.942      0.000
-0.222      -0.154
concavity_mean              -0.1206      0.016     -7.498      0.000
-0.152      -0.089
fractal_dimension_mean       0.0111      0.011      0.971      0.332
-0.011       0.034
radius_se                    0.0286      0.023      1.260      0.208
-0.016       0.073
texture_se                   0.0049      0.005      0.973      0.331
-0.005       0.015
perimeter_se                -0.0670      0.023     -2.916      0.004
-0.112      -0.022
area_se                     -0.0026      0.017     -0.149      0.882
-0.036       0.031
smoothness_se                0.0094      0.005      1.730      0.084
-0.001       0.020
compactness_se              -0.0126      0.010     -1.330      0.184
-0.031       0.006
concavity_se                 0.0402      0.010      3.951      0.000
0.020       0.060
concave points_se            0.0134      0.009      1.408      0.160
-0.005       0.032
symmetry_se                 -0.0027      0.006     -0.445      0.656
-0.015       0.009
radius_worst                 0.7726      0.065     11.849      0.000
0.644       0.901
texture_worst               -0.0069      0.005     -1.419      0.157
-0.016       0.003
perimeter_worst             -0.3757      0.055     -6.850      0.000
-0.483      -0.268
area_worst                  -0.2736      0.047     -5.792      0.000
-0.366      -0.181
smoothness_worst            -0.0311      0.009     -3.546      0.000
-0.048      -0.014
compactness_worst            0.0586      0.016      3.601      0.000
0.027       0.091
concavity_worst              0.0005      0.015      0.032      0.974
```

```
-0.028       0.029
concave points_worst         0.0111      0.014      0.785      0.433
-0.017       0.039
symmetry_worst               0.0016      0.006      0.251      0.802
-0.011       0.014
fractal_dimension_worst    -0.0164      0.011     -1.520      0.129
-0.038       0.005
==============================================================================
========
Omnibus:                       55.665   Durbin-Watson:
1.960
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
430.695
Skew:                           0.009   Prob(JB):
2.99e-94
Kurtosis:                       7.766   Cond. No.
119.
==============================================================================
========

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

Eliminating `radius_se` with `P>|t|` = 0.350

```
x_backward.drop(["radius_se"], inplace=True, axis=1)

model_backward = sm.OLS(y_train.values, x_backward).fit()
print(model_backward.summary())

                            OLS Regression Results

==============================================================================
========
Dep. Variable:                         y   R-squared:
1.000
Model:                               OLS   Adj. R-squared:
1.000
Method:                    Least Squares   F-statistic:
6.486e+04
Date:                   Tue, 12 Sep 2023   Prob (F-statistic):
0.00
Time:                           05:13:28   Log-Likelihood:
648.22
No. Observations:                   455   AIC:
-1246.
Df Residuals:                       430   BIC:
-1143.
```

```
Df Model:                           24

Covariance Type:             nonrobust

==============================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const                        14.0186      0.003   4993.397      0.000      14.013      14.024
perimeter_mean               3.2418      0.061     53.246      0.000       3.122       3.361
area_mean                    0.2954      0.049      6.069      0.000       0.200       0.391
smoothness_mean              0.0322      0.008      4.244      0.000       0.017       0.047
compactness_mean            -0.1877      0.017    -10.905      0.000      -0.222      -0.154
concavity_mean              -0.1162      0.016     -7.396      0.000      -0.147      -0.085
fractal_dimension_mean       0.0102      0.011      0.893      0.372      -0.012       0.033
texture_se                   0.0053      0.005      1.038      0.300      -0.005       0.015
perimeter_se                -0.0471      0.017     -2.823      0.005      -0.080      -0.014
area_se                      0.0052      0.016      0.319      0.750      -0.027       0.037
smoothness_se                0.0105      0.005      1.957      0.051   -4.35e-05       0.021
compactness_se              -0.0120      0.010     -1.265      0.207      -0.031       0.007
concavity_se                 0.0402      0.010      3.953      0.000       0.020       0.060
concave points_se            0.0135      0.009      1.420      0.156      -0.005       0.032
symmetry_se                 -0.0025      0.006     -0.416      0.678      -0.014       0.009
radius_worst                 0.8284      0.048     17.283      0.000       0.734       0.923
texture_worst               -0.0072      0.005     -1.495      0.136      -0.017       0.002
perimeter_worst             -0.4069      0.049     -8.314      0.000      -0.503      -0.311
area_worst                  -0.2870      0.046     -6.232      0.000      -0.377      -0.196
smoothness_worst            -0.0323      0.009     -3.706      0.000
```

```
-0.049      -0.015
compactness_worst           0.0587      0.016      3.601      0.000
0.027       0.091
concavity_worst            -0.0015      0.015     -0.105      0.916
-0.030       0.027
concave points_worst        0.0099      0.014      0.699      0.485
-0.018       0.038
symmetry_worst              0.0013      0.006      0.206      0.837
-0.011       0.014
fractal_dimension_worst    -0.0161      0.011     -1.491      0.137
-0.037       0.005
==============================================================================
========
Omnibus:                       57.150   Durbin-Watson:
1.960
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
460.430
Skew:                          -0.015   Prob(JB):
1.04e-100
Kurtosis:                       7.928   Cond. No.
112.
==============================================================================
========

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

Eliminating `concavity_worst` with `P>|t|` = 0.405

```
x_backward.drop(["concavity_worst"], inplace=True, axis=1)

model_backward = sm.OLS(y_train.values, x_backward).fit()
print(model_backward.summary())
```

```
                            OLS Regression Results

==============================================================================
========
Dep. Variable:                       y   R-squared:
1.000
Model:                             OLS   Adj. R-squared:
1.000
Method:                  Least Squares   F-statistic:
6.784e+04
Date:                 Tue, 12 Sep 2023   Prob (F-statistic):
0.00
Time:                        05:13:28   Log-Likelihood:
648.21
```

```
No. Observations:                 455   AIC:
-1248.
Df Residuals:                     431   BIC:
-1150.
Df Model:                          23

Covariance Type:            nonrobust

==============================================================================
                                 coef    std err          t      P>|t|
[0.025      0.975]
------------------------------------------------------------------------------
const                        14.0186      0.003   4999.135      0.000
14.013      14.024
perimeter_mean                3.2410      0.060     53.658      0.000
3.122       3.360
area_mean                     0.2961      0.048      6.150      0.000
0.201       0.391
smoothness_mean               0.0322      0.008      4.249      0.000
0.017       0.047
compactness_mean             -0.1873      0.017    -11.189      0.000
-0.220      -0.154
concavity_mean               -0.1171      0.014     -8.667      0.000
-0.144      -0.091
fractal_dimension_mean        0.0104      0.011      0.919      0.359
-0.012       0.033
texture_se                    0.0053      0.005      1.056      0.292
-0.005       0.015
perimeter_se                 -0.0473      0.017     -2.858      0.004
-0.080      -0.015
area_se                       0.0054      0.016      0.341      0.733
-0.026       0.037
smoothness_se                 0.0106      0.005      1.973      0.049
3.77e-05       0.021
compactness_se               -0.0118      0.009     -1.268      0.206
-0.030       0.007
concavity_se                  0.0397      0.009      4.355      0.000
0.022       0.058
concave points_se             0.0137      0.009      1.467      0.143
-0.005       0.032
symmetry_se                  -0.0026      0.006     -0.421      0.674
-0.014       0.009
radius_worst                  0.8290      0.048     17.451      0.000
0.736       0.922
texture_worst                -0.0073      0.005     -1.508      0.132
-0.017       0.002
perimeter_worst              -0.4066      0.049     -8.335      0.000
```

```
-0.502      -0.311
area_worst                -0.2878      0.045     -6.349      0.000
-0.377      -0.199
smoothness_worst          -0.0324      0.009     -3.733      0.000
-0.049      -0.015
compactness_worst          0.0579      0.014      4.055      0.000
0.030       0.086
concave points_worst       0.0096      0.014      0.694      0.488
-0.018       0.037
symmetry_worst             0.0013      0.006      0.213      0.831
-0.011       0.014
fractal_dimension_worst   -0.0162      0.011     -1.511      0.132
-0.037       0.005
==============================================================================
========
Omnibus:                       57.087   Durbin-Watson:
1.961
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
458.983
Skew:                          -0.017   Prob(JB):
2.15e-100
Kurtosis:                       7.920   Cond. No.
106.
==============================================================================
========

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

Eliminating `symmetry_se` with `P>|t|` = 0.326

```
x_backward.drop(["symmetry_se"], inplace=True, axis=1)

model_backward = sm.OLS(y_train.values, x_backward).fit()
print(model_backward.summary())
```

```
                            OLS Regression Results

==============================================================================
========
Dep. Variable:                         y   R-squared:
1.000
Model:                               OLS   Adj. R-squared:
1.000
Method:                    Least Squares   F-statistic:
7.106e+04
Date:                   Tue, 12 Sep 2023   Prob (F-statistic):
0.00
```

```
Time:                         05:13:28   Log-Likelihood:
648.12
No. Observations:                455   AIC:
-1250.
Df Residuals:                    432   BIC:
-1155.
Df Model:                         22

Covariance Type:            nonrobust

==============================================================================
======================
                                 coef    std err          t      P>|t|
[0.025      0.975]
------------------------------------------------------------------------------
----------------------
const                        14.0186      0.003   5003.904      0.000
14.013      14.024
perimeter_mean                3.2391      0.060     53.837      0.000
3.121       3.357
area_mean                     0.2966      0.048      6.168      0.000
0.202       0.391
smoothness_mean               0.0321      0.008      4.239      0.000
0.017       0.047
compactness_mean             -0.1874      0.017    -11.208      0.000
-0.220      -0.155
concavity_mean               -0.1173      0.013     -8.699      0.000
-0.144      -0.091
fractal_dimension_mean        0.0101      0.011      0.895      0.372
-0.012       0.032
texture_se                    0.0049      0.005      0.995      0.320
-0.005       0.015
perimeter_se                 -0.0490      0.016     -3.065      0.002
-0.080      -0.018
area_se                       0.0064      0.016      0.410      0.682
-0.024       0.037
smoothness_se                 0.0100      0.005      1.929      0.054
-0.000       0.020
compactness_se               -0.0125      0.009     -1.357      0.176
-0.031       0.006
concavity_se                  0.0399      0.009      4.383      0.000
0.022       0.058
concave points_se             0.0137      0.009      1.469      0.143
-0.005       0.032
radius_worst                  0.8293      0.047     17.475      0.000
0.736       0.923
texture_worst                -0.0070      0.005     -1.467      0.143
-0.016       0.002
perimeter_worst              -0.4034      0.048     -8.379      0.000
```

```
-0.498      -0.309
area_worst                -0.2892      0.045     -6.402      0.000
-0.378      -0.200
smoothness_worst          -0.0317      0.009     -3.722      0.000
-0.048      -0.015
compactness_worst          0.0582      0.014      4.092      0.000
0.030       0.086
concave points_worst       0.0098      0.014      0.715      0.475
-0.017       0.037
symmetry_worst            -0.0007      0.004     -0.173      0.863
-0.009       0.007
fractal_dimension_worst   -0.0156      0.011     -1.467      0.143
-0.036       0.005
==============================================================================
========
Omnibus:                       58.847   Durbin-Watson:
1.959
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
494.862
Skew:                          -0.037   Prob(JB):
3.48e-108
Kurtosis:                       8.109   Cond. No.
106.
==============================================================================
========

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

Eliminating `symmetry_worst` with `P>|t|` = 0.621

```
x_backward.drop(["symmetry_worst"], inplace=True, axis=1)

model_backward = sm.OLS(y_train.values, x_backward).fit()
print(model_backward.summary())

                            OLS Regression Results

==============================================================================
========
Dep. Variable:                         y   R-squared:
1.000
Model:                               OLS   Adj. R-squared:
1.000
Method:                    Least Squares   F-statistic:
7.461e+04
Date:                   Tue, 12 Sep 2023   Prob (F-statistic):
0.00
```

```
Time:                         05:13:28   Log-Likelihood:
648.10
No. Observations:                455   AIC:
-1252.
Df Residuals:                    433   BIC:
-1162.
Df Model:                         21

Covariance Type:            nonrobust

==============================================================================
======================
                                 coef    std err          t      P>|t|
[0.025      0.975]
------------------------------------------------------------------------------
----------------------
const                        14.0186      0.003   5009.519      0.000
14.013      14.024
perimeter_mean                3.2403      0.060     54.267      0.000
3.123       3.358
area_mean                     0.2959      0.048      6.182      0.000
0.202       0.390
smoothness_mean               0.0321      0.008      4.249      0.000
0.017       0.047
compactness_mean             -0.1876      0.017    -11.250      0.000
-0.220      -0.155
concavity_mean               -0.1172      0.013     -8.707      0.000
-0.144      -0.091
fractal_dimension_mean        0.0100      0.011      0.887      0.375
-0.012       0.032
texture_se                    0.0049      0.005      0.994      0.321
-0.005       0.015
perimeter_se                 -0.0492      0.016     -3.085      0.002
-0.081      -0.018
area_se                       0.0066      0.016      0.420      0.675
-0.024       0.037
smoothness_se                 0.0101      0.005      1.947      0.052
-9.35e-05       0.020
compactness_se               -0.0125      0.009     -1.355      0.176
-0.031       0.006
concavity_se                  0.0398      0.009      4.385      0.000
0.022       0.058
concave points_se             0.0139      0.009      1.507      0.132
-0.004       0.032
radius_worst                  0.8278      0.047     17.733      0.000
0.736       0.920
texture_worst                -0.0070      0.005     -1.466      0.143
-0.016       0.002
perimeter_worst              -0.4030      0.048     -8.389      0.000
```

```
-0.497      -0.309
area_worst                -0.2884      0.045     -6.424      0.000
-0.377      -0.200
smoothness_worst          -0.0318      0.008     -3.746      0.000
-0.048      -0.015
compactness_worst          0.0579      0.014      4.121      0.000
0.030       0.085
concave points_worst       0.0096      0.014      0.700      0.484
-0.017       0.036
fractal_dimension_worst   -0.0154      0.011     -1.458      0.146
-0.036       0.005
==============================================================================
Omnibus:                       59.319   Durbin-Watson:
1.959
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
503.482
Skew:                          -0.049   Prob(JB):
4.68e-110
Kurtosis:                       8.152   Cond. No.
104.
==============================================================================

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

Eliminating `fractal_dimension_mean` with `P>|t|` = 0.240

```
x_backward.drop(["fractal_dimension_mean"], inplace=True, axis=1)

model_backward = sm.OLS(y_train.values, x_backward).fit()
print(model_backward.summary())
```

```
                            OLS Regression Results

==============================================================================
Dep. Variable:                      y   R-squared:
1.000
Model:                            OLS   Adj. R-squared:
1.000
Method:                 Least Squares   F-statistic:
7.838e+04
Date:                Tue, 12 Sep 2023   Prob (F-statistic):
0.00
Time:                        05:13:29   Log-Likelihood:
647.69
```

```
No. Observations:                 455   AIC:
-1253.
Df Residuals:                     434   BIC:
-1167.
Df Model:                          20

Covariance Type:            nonrobust

==============================================================================
                                 coef    std err          t      P>|t|
[0.025      0.975]
------------------------------------------------------------------------------
const                        14.0186      0.003   5010.747      0.000
14.013      14.024
perimeter_mean                3.2293      0.058     55.285      0.000
3.115       3.344
area_mean                     0.3031      0.047      6.428      0.000
0.210       0.396
smoothness_mean               0.0335      0.007      4.524      0.000
0.019       0.048
compactness_mean             -0.1783      0.013    -13.791      0.000
-0.204      -0.153
concavity_mean               -0.1188      0.013     -8.902      0.000
-0.145      -0.093
texture_se                    0.0049      0.005      0.987      0.324
-0.005       0.015
perimeter_se                 -0.0476      0.016     -3.004      0.003
-0.079      -0.016
area_se                       0.0043      0.015      0.280      0.780
-0.026       0.035
smoothness_se                 0.0102      0.005      1.962      0.050
-1.9e-05       0.020
compactness_se               -0.0122      0.009     -1.326      0.186
-0.030       0.006
concavity_se                  0.0412      0.009      4.611      0.000
0.024       0.059
concave points_se             0.0132      0.009      1.441      0.150
-0.005       0.031
radius_worst                  0.8249      0.047     17.719      0.000
0.733       0.916
texture_worst                -0.0072      0.005     -1.511      0.132
-0.017       0.002
perimeter_worst              -0.4013      0.048     -8.363      0.000
-0.496      -0.307
area_worst                   -0.2869      0.045     -6.397      0.000
-0.375      -0.199
smoothness_worst             -0.0324      0.008     -3.838      0.000
```

```
-0.049      -0.016
compactness_worst           0.0512      0.012      4.307      0.000
0.028       0.075
concave points_worst        0.0095      0.014      0.697      0.486
-0.017       0.036
fractal_dimension_worst    -0.0085      0.007     -1.191      0.234
-0.022       0.006
==============================================================================
========
Omnibus:                       58.726   Durbin-Watson:
1.962
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
490.761
Skew:                          -0.047   Prob(JB):
2.71e-107
Kurtosis:                       8.087   Cond. No.
103.
==============================================================================
========

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

Eliminating `fractal_dimension_worst` with `P>|t|` = 0.254

```
x_backward.drop(["fractal_dimension_worst"], inplace=True, axis=1)

model_backward = sm.OLS(y_train.values, x_backward).fit()
print(model_backward.summary())

                            OLS Regression Results

==============================================================================
========
Dep. Variable:                        y   R-squared:
1.000
Model:                              OLS   Adj. R-squared:
1.000
Method:                   Least Squares   F-statistic:
8.242e+04
Date:                  Tue, 12 Sep 2023   Prob (F-statistic):
0.00
Time:                          05:13:29   Log-Likelihood:
646.95
No. Observations:                 455   AIC:
-1254.
Df Residuals:                     435   BIC:
-1171.
```

```
Df Model:                           19

Covariance Type:             nonrobust

==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const                   14.0186      0.003   5008.338      0.000      14.013      14.024
perimeter_mean           3.2372      0.058     55.754      0.000       3.123       3.351
area_mean                0.2978      0.047      6.341      0.000       0.205       0.390
smoothness_mean          0.0342      0.007      4.633      0.000       0.020       0.049
compactness_mean        -0.1814      0.013    -14.342      0.000      -0.206      -0.157
concavity_mean          -0.1189      0.013     -8.905      0.000      -0.145      -0.093
texture_se               0.0049      0.005      0.991      0.322      -0.005       0.015
perimeter_se            -0.0487      0.016     -3.083      0.002      -0.080      -0.018
area_se                  0.0060      0.015      0.389      0.697      -0.024       0.036
smoothness_se            0.0106      0.005      2.053      0.041       0.000       0.021
compactness_se          -0.0119      0.009     -1.299      0.195      -0.030       0.006
concavity_se             0.0404      0.009      4.534      0.000       0.023       0.058
concave points_se        0.0137      0.009      1.489      0.137      -0.004       0.032
radius_worst             0.8144      0.046     17.809      0.000       0.725       0.904
texture_worst           -0.0070      0.005     -1.480      0.140      -0.016       0.002
perimeter_worst         -0.3874      0.047     -8.319      0.000      -0.479      -0.296
area_worst              -0.2890      0.045     -6.446      0.000      -0.377      -0.201
smoothness_worst        -0.0341      0.008     -4.085      0.000      -0.050      -0.018
compactness_worst        0.0446      0.011      4.238      0.000       0.024       0.065
concave points_worst     0.0096      0.014      0.705      0.481
```

```
-0.017       0.036
==============================================================================
========
Omnibus:                       57.124   Durbin-Watson:
1.964
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
459.842
Skew:                          -0.016   Prob(JB):
1.40e-100
Kurtosis:                       7.925   Cond. No.
102.
==============================================================================
========

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

Eliminating `texture_se` with `P>|t|` = 0.085

```
x_backward.drop(["texture_se"], inplace=True, axis=1)

model_backward = sm.OLS(y_train.values, x_backward).fit()
print(model_backward.summary())

                            OLS Regression Results

==============================================================================
========
Dep. Variable:                      y   R-squared:
1.000
Model:                            OLS   Adj. R-squared:
1.000
Method:                 Least Squares   F-statistic:
8.700e+04
Date:                Tue, 12 Sep 2023   Prob (F-statistic):
0.00
Time:                        05:13:29   Log-Likelihood:
646.43
No. Observations:                 455   AIC:
-1255.
Df Residuals:                     436   BIC:
-1177.
Df Model:                          18

Covariance Type:            nonrobust

==============================================================================
==================
                           coef    std err          t      P>|t|
```

```
[0.025      0.975]
------------------------------------------------------------------------------
------------------
const                    14.0186      0.003   5008.438      0.000
14.013      14.024
perimeter_mean            3.2449      0.058     56.389      0.000
3.132       3.358
area_mean                 0.2925      0.047      6.269      0.000
0.201       0.384
smoothness_mean           0.0360      0.007      5.036      0.000
0.022       0.050
compactness_mean         -0.1826      0.013    -14.488      0.000
-0.207      -0.158
concavity_mean           -0.1178      0.013     -8.853      0.000
-0.144      -0.092
perimeter_se             -0.0445      0.015     -2.924      0.004
-0.074      -0.015
area_se                   0.0027      0.015      0.180      0.857
-0.027       0.032
smoothness_se             0.0116      0.005      2.295      0.022
0.002       0.022
compactness_se           -0.0108      0.009     -1.188      0.235
-0.029       0.007
concavity_se              0.0401      0.009      4.501      0.000
0.023       0.058
concave points_se         0.0138      0.009      1.506      0.133
-0.004       0.032
radius_worst              0.8110      0.046     17.785      0.000
0.721       0.901
texture_worst            -0.0037      0.003     -1.103      0.271
-0.010       0.003
perimeter_worst          -0.3947      0.046     -8.584      0.000
-0.485      -0.304
area_worst               -0.2818      0.044     -6.370      0.000
-0.369      -0.195
smoothness_worst         -0.0355      0.008     -4.325      0.000
-0.052      -0.019
compactness_worst         0.0438      0.010      4.171      0.000
0.023       0.064
concave points_worst      0.0086      0.014      0.631      0.528
-0.018       0.035
==============================================================================
========
Omnibus:                       57.448   Durbin-Watson:
1.970
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
466.345
Skew:                          -0.020   Prob(JB):
5.43e-102
```

```
Kurtosis:                          7.960   Cond. No.
101.
==============================================================================
========

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

Eliminating `texture_worst` with `P>|t|` = 0.090

```
x_backward.drop(["texture_worst"], inplace=True, axis=1)

model_backward = sm.OLS(y_train.values, x_backward).fit()
print(model_backward.summary())

                            OLS Regression Results

==============================================================================
========
Dep. Variable:                        y   R-squared:
1.000
Model:                              OLS   Adj. R-squared:
1.000
Method:                   Least Squares   F-statistic:
9.208e+04
Date:                  Tue, 12 Sep 2023   Prob (F-statistic):
0.00
Time:                          05:13:29   Log-Likelihood:
645.80
No. Observations:                   455   AIC:
-1256.
Df Residuals:                       437   BIC:
-1181.
Df Model:                            17

Covariance Type:             nonrobust

==============================================================================
==================
                             coef    std err          t      P>|t|
[0.025      0.975]
------------------------------------------------------------------------------
------------------
const                     14.0186      0.003   5007.202      0.000
14.013      14.024
perimeter_mean             3.2446      0.058     56.370      0.000
3.131       3.358
area_mean                  0.2942      0.047      6.308      0.000
0.203       0.386
```

```
smoothness_mean           0.0376      0.007      5.352      0.000
0.024       0.051
compactness_mean         -0.1814      0.013    -14.443      0.000
-0.206      -0.157
concavity_mean           -0.1194      0.013     -9.032      0.000
-0.145      -0.093
perimeter_se             -0.0458      0.015     -3.017      0.003
-0.076      -0.016
area_se                   0.0040      0.015      0.267      0.790
-0.025       0.034
smoothness_se             0.0114      0.005      2.251      0.025
0.001       0.021
compactness_se           -0.0107      0.009     -1.173      0.242
-0.029       0.007
concavity_se              0.0407      0.009      4.567      0.000
0.023       0.058
concave points_se         0.0140      0.009      1.521      0.129
-0.004       0.032
radius_worst              0.8086      0.046     17.748      0.000
0.719       0.898
perimeter_worst          -0.3929      0.046     -8.548      0.000
-0.483      -0.303
area_worst               -0.2834      0.044     -6.407      0.000
-0.370      -0.196
smoothness_worst         -0.0368      0.008     -4.515      0.000
-0.053      -0.021
compactness_worst         0.0421      0.010      4.052      0.000
0.022       0.062
concave points_worst      0.0092      0.014      0.676      0.499
-0.018       0.036
==============================================================================
Omnibus:                       58.604   Durbin-Watson:
1.962
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
490.606
Skew:                          -0.026   Prob(JB):
2.93e-107
Kurtosis:                       8.087   Cond. No.
100.
==============================================================================

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

Eliminating `concave points_worst` with `P>|t|` = 0.074

```
x_backward.drop(["concave points_worst"], inplace=True, axis=1)

model_backward = sm.OLS(y_train.values, x_backward).fit()
print(model_backward.summary())
```

```
                            OLS Regression Results

==============================================================================
Dep. Variable:                      y   R-squared:
1.000
Model:                            OLS   Adj. R-squared:
1.000
Method:                 Least Squares   F-statistic:
9.795e+04
Date:                Tue, 12 Sep 2023   Prob (F-statistic):
0.00
Time:                        05:13:29   Log-Likelihood:
645.56
No. Observations:                 455   AIC:
-1257.
Df Residuals:                     438   BIC:
-1187.
Df Model:                          16

Covariance Type:            nonrobust

==============================================================================
                 coef    std err          t      P>|t|
[0.025      0.975]
------------------------------------------------------------------------------
const                14.0186      0.003   5010.310      0.000
14.013      14.024
perimeter_mean        3.2346      0.056     58.196      0.000
3.125       3.344
area_mean             0.3018      0.045      6.669      0.000
0.213       0.391
smoothness_mean       0.0369      0.007      5.314      0.000
0.023       0.051
compactness_mean     -0.1813      0.013    -14.444      0.000      -
0.206      -0.157
concavity_mean       -0.1157      0.012     -9.612      0.000      -
0.139      -0.092
perimeter_se         -0.0491      0.014     -3.431      0.001      -
0.077      -0.021
area_se               0.0061      0.015      0.418      0.676      -
0.023       0.035
smoothness_se         0.0104      0.005      2.149      0.032
```

```
0.001       0.020
compactness_se       -0.0120      0.009     -1.350      0.178      -
0.030       0.005
concavity_se          0.0386      0.008      4.620      0.000
0.022       0.055
concave points_se     0.0185      0.006      2.970      0.003
0.006       0.031
radius_worst          0.8158      0.044     18.422      0.000
0.729       0.903
perimeter_worst      -0.3865      0.045     -8.598      0.000      -
0.475      -0.298
area_worst           -0.2918      0.042     -6.882      0.000      -
0.375      -0.208
smoothness_worst     -0.0349      0.008     -4.556      0.000      -
0.050      -0.020
compactness_worst     0.0447      0.010      4.635      0.000
0.026       0.064
==============================================================================
========
Omnibus:                       58.438   Durbin-Watson:
1.968
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
484.722
Skew:                          -0.046   Prob(JB):
5.55e-106
Kurtosis:                       8.056   Cond. No.
91.8
==============================================================================
========

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

Eliminating `compactness_se` with `P>|t|` = 0.099

```
x_backward.drop(["compactness_se"], inplace=True, axis=1)

model_backward = sm.OLS(y_train.values, x_backward).fit()
print(model_backward.summary())

                            OLS Regression Results

==============================================================================
========
Dep. Variable:                       y   R-squared:
1.000
Model:                             OLS   Adj. R-squared:
1.000
```

```
Method:                  Least Squares   F-statistic:
1.043e+05
Date:                 Tue, 12 Sep 2023   Prob (F-statistic):
0.00
Time:                         05:13:29   Log-Likelihood:
644.62
No. Observations:                 455   AIC:
-1257.
Df Residuals:                     439   BIC:
-1191.
Df Model:                          15

Covariance Type:            nonrobust

==============================================================================
================
                     coef    std err          t      P>|t|
[0.025      0.975]
------------------------------------------------------------------------------
---------------
const                14.0186      0.003   5005.621      0.000
14.013      14.024
perimeter_mean        3.2278      0.055     58.255      0.000
3.119       3.337
area_mean             0.3059      0.045      6.771      0.000
0.217       0.395
smoothness_mean       0.0376      0.007      5.422      0.000
0.024       0.051
compactness_mean     -0.1870      0.012    -15.807      0.000       -
0.210      -0.164
concavity_mean       -0.1104      0.011     -9.703      0.000       -
0.133      -0.088
perimeter_se         -0.0504      0.014     -3.526      0.000       -
0.079      -0.022
area_se               0.0063      0.015      0.427      0.670       -
0.023       0.035
smoothness_se         0.0072      0.004      1.705      0.089       -
0.001       0.016
concavity_se          0.0329      0.007      4.552      0.000
0.019       0.047
concave points_se     0.0174      0.006      2.808      0.005
0.005       0.030
radius_worst          0.8190      0.044     18.505      0.000
0.732       0.906
perimeter_worst      -0.3808      0.045     -8.500      0.000       -
0.469      -0.293
area_worst           -0.2976      0.042     -7.048      0.000       -
0.381      -0.215
smoothness_worst     -0.0317      0.007     -4.348      0.000       -
```

```
0.046      -0.017
compactness_worst     0.0382      0.008      4.566      0.000
0.022       0.055
==============================================================================
========
Omnibus:                       55.551   Durbin-Watson:
1.958
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
428.175
Skew:                          -0.017   Prob(JB):
1.05e-93
Kurtosis:                       7.752   Cond. No.
89.8
==============================================================================
========

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

Eliminating `smoothness_se` with `P>|t|` = 0.034

```
x_backward.drop(["smoothness_se"], inplace=True, axis=1)

model_backward = sm.OLS(y_train.values, x_backward).fit()
print(model_backward.summary())
```

```
                            OLS Regression Results

==============================================================================
========
Dep. Variable:                        y   R-squared:
1.000
Model:                              OLS   Adj. R-squared:
1.000
Method:                   Least Squares   F-statistic:
1.113e+05
Date:                  Tue, 12 Sep 2023   Prob (F-statistic):
0.00
Time:                          05:13:29   Log-Likelihood:
643.12
No. Observations:                   455   AIC:
-1256.
Df Residuals:                       440   BIC:
-1194.
Df Model:                            14

Covariance Type:             nonrobust

==============================================================================
```

```
===============
                         coef    std err          t      P>|t|
[0.025      0.975]
------------------------------------------------------------------------
---------------
const                 14.0186      0.003   4994.808      0.000
14.013      14.024
perimeter_mean         3.2330      0.055     58.308      0.000
3.124       3.342
area_mean              0.3087      0.045      6.821      0.000
0.220       0.398
smoothness_mean        0.0350      0.007      5.165      0.000
0.022       0.048
compactness_mean      -0.1844      0.012    -15.683      0.000      -
0.208      -0.161
concavity_mean        -0.1112      0.011     -9.763      0.000      -
0.134      -0.089
perimeter_se          -0.0455      0.014     -3.243      0.001      -
0.073      -0.018
area_se                0.0035      0.015      0.239      0.812      -
0.025       0.032
concavity_se           0.0334      0.007      4.613      0.000
0.019       0.048
concave points_se      0.0194      0.006      3.181      0.002
0.007       0.031
radius_worst           0.8128      0.044     18.388      0.000
0.726       0.900
perimeter_worst       -0.3944      0.044     -8.928      0.000      -
0.481      -0.308
area_worst            -0.2891      0.042     -6.880      0.000      -
0.372      -0.207
smoothness_worst      -0.0259      0.006     -4.007      0.000      -
0.039      -0.013
compactness_worst      0.0352      0.008      4.297      0.000
0.019       0.051
==============================================================================
========
Omnibus:                       57.082   Durbin-Watson:
1.942
Prob(Omnibus):                  0.000   Jarque-Bera (JB):
459.159
Skew:                          -0.011   Prob(JB):
1.97e-100
Kurtosis:                       7.921   Cond. No.
89.5
==============================================================================
========

Notes:
```

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

As we can see, the multiple regression model was much simpler, reducing the independent variables to 14, which reduces the complexity of the model, reduces the computational costs for modeling and in this case we see that the accuracy of the model is maintained even with slightly less than half the number of variables.

6. Comparison between real data and prediction. Analysis of results.

```
x_test_std
```

```
array([[ 0.53466649,  0.12542914,  0.00889231, ...,  0.14135664,
         0.19736677, -0.8526808 ],
       [-0.46714672, -0.4389997 , -0.43267045, ..., -0.22726897,
         0.5979247 , -0.59217342],
       [-1.23652067, -0.35245119, -0.32547846, ..., -1.19823406,
        -1.26547076, -0.99585221],
       ...,
       [ 0.62299341, -0.16817325, -0.23904248, ..., -0.78996468,
        -0.72231422, -0.75687436],
       [-0.04875605, -0.90859787, -0.84921227, ..., -0.52306068,
         0.06918824, -0.45384615],
       [ 1.07160118, -0.11102638, -0.1565871 , ..., -0.80271829,
        -0.37462994, -1.13525595]])
```

```
x_test_std_df = pd.DataFrame(x_test_std, columns=x_train_std.columns)
```

```
x_test_std_df
```

```
     texture_mean  perimeter_mean  area_mean  smoothness_mean  \
0        0.534666        0.125429   0.008892        -0.140656
1       -0.467147       -0.439000  -0.432670        -0.737818
2       -1.236521       -0.352451  -0.325478        -0.898032
3        0.590452       -1.309454  -1.063028         0.550138
4        0.646237       -1.333887  -1.083499        -0.282144
..            ...             ...        ...              ...
109     -0.078973        1.527184   1.483991         0.862244
110      0.762457       -0.540042  -0.507733        -1.009003
111      0.622993       -0.168173  -0.239042        -1.172685
112     -0.048756       -0.908598  -0.849212         0.028575
113      1.071601       -0.111026  -0.156587        -1.112345

     compactness_mean  concavity_mean  concave points_mean
symmetry_mean  \
0           -0.079782        0.405231             0.036943
0.956099
1           -0.941403       -0.726757            -0.668267      -
0.040778
2           -1.264726       -1.039893            -1.035711      -
1.290487
```

```
3           -0.802346       -0.841403             -0.936754
0.775505
4           -0.908022       -0.598264             -0.930141      -
0.427249
..                ...             ...                   ...
...
109          0.784892        1.717608              1.889109
0.045907
110         -0.985276       -0.882945             -0.935696      -
0.961806
111         -0.401769       -0.588782             -0.702654      -
0.633126
112          0.210732       -0.299985             -0.570394
0.403482
113         -0.678739       -0.510813             -0.649221      -
0.134687

     fractal_dimension_mean  radius_se  ...  radius_worst
texture_worst  \
0                -0.912253   0.078854  ...      0.017629
0.606267
1                -0.649545  -0.813272  ...     -0.370870      -
0.546137
2                -1.075409  -0.002474  ...     -0.289846      -
1.317112
3                 0.422029   0.113658  ...     -1.095929
0.929265
4                 0.387462  -0.344125  ...     -1.304722      -
0.075437
..                     ...        ...  ...           ...             .
..
109              -0.249951   0.595591  ...      1.627718
0.791301
110              -0.442143  -0.619007  ...     -0.406188
1.050998
111              -0.603916  -0.213786  ...     -0.275303
0.734493
112               0.269935   0.185753  ...     -0.881943      -
0.417912
113              -1.198467   0.129640  ...     -0.231675
0.133944

     perimeter_worst  area_worst  smoothness_worst  compactness_worst
\
0           0.066418   -0.099769         -0.090726          -0.347143

1          -0.435601   -0.407070         -0.446279          -0.637571

2          -0.365593   -0.338917         -1.363696          -1.227476
```

```
3          -1.120417   -0.887293          0.874976          -0.842579

4          -1.296034   -0.998720         -0.981805          -1.072391

..               ...         ...               ...                ...

109         1.427673    1.531517          0.822301           0.759523

110        -0.402093   -0.425641         -0.310203          -0.455974

111        -0.254898   -0.358014         -1.411981          -0.502797

112        -0.834703   -0.772712         -0.498954          -0.006095

113        -0.244726   -0.280050         -1.711348          -0.685027


     concavity_worst  concave points_worst  symmetry_worst  \
0           0.296294              0.141357        0.197367
1          -0.387396             -0.227269        0.597925
2          -1.184535             -1.198234       -1.265471
3          -0.982655             -1.161663       -0.039764
4          -0.896048             -1.346975       -0.739939
..               ...                   ...             ...
109         1.506078              1.424401        0.591516
110        -0.694635             -0.767531        0.325545
111        -0.590352             -0.789965       -0.722314
112        -0.373834             -0.523061        0.069188
113        -0.634310             -0.802718       -0.374630

     fractal_dimension_worst
0                  -0.852681
1                  -0.592173
2                  -0.995852
3                  -0.067929
4                  -0.558264
..                       ...
109                 0.311529
110                -0.122291
111                -0.756874
112                -0.453846
113                -1.135256

[114 rows x 29 columns]

x_test_std_df_cons = sm.add_constant(x_test_std_df)

x_test_std_df_cons.drop(["texture_mean", "concave points_mean",
"fractal_dimension_se", "symmetry_mean", "radius_se",
"concavity_worst", "symmetry_se", "symmetry_worst",
"fractal_dimension_mean",
```

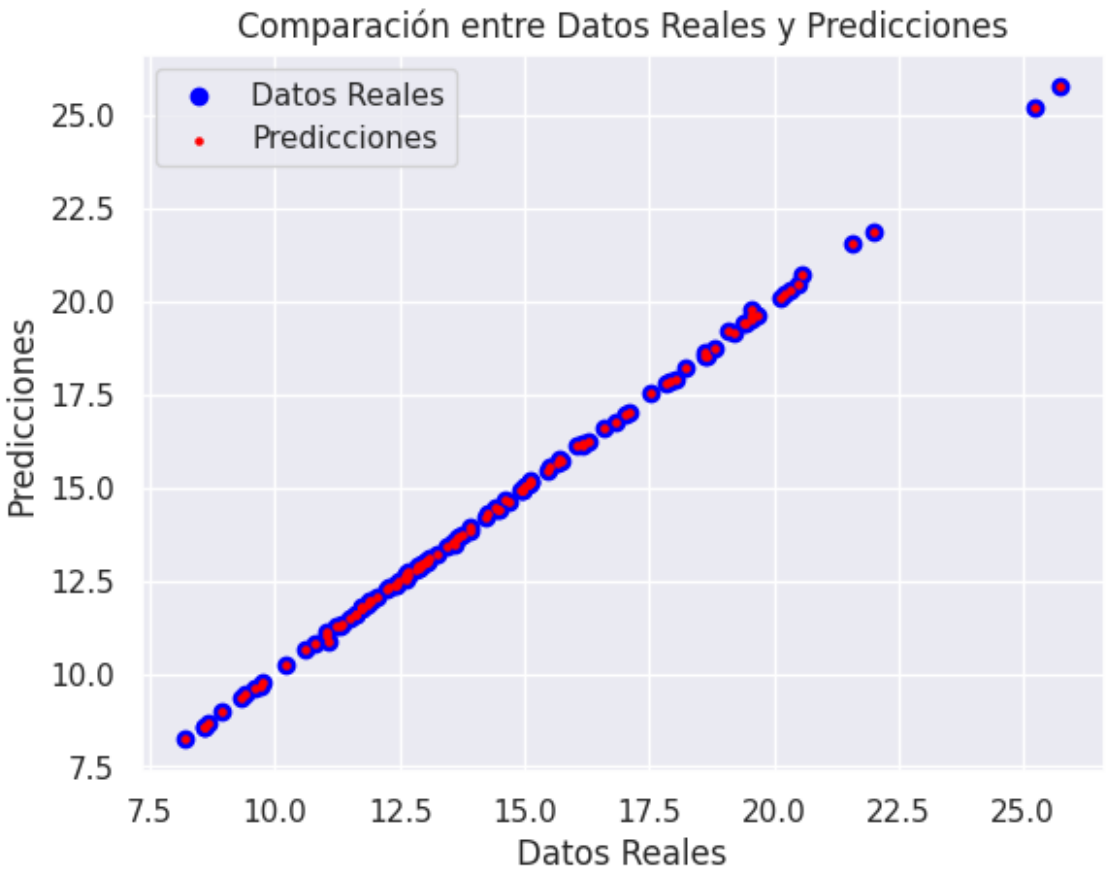
```
"fractal_dimension_worst", "texture_se", "texture_worst", "concave
points_worst",
"compactness_se", "smoothness_se"], inplace=True, axis=1)

y_hat = model_backward.predict(x_test_std_df_cons)

y_hat
```

```
0      18.876613
1      18.383824
2      15.559358
3      11.537208
4       9.656352
         ...
450    13.851986
451    12.090348
452    11.836651
453    11.360343
454    14.019527
Length: 455, dtype: float64
```

```
y_test
```

```
86     14.480
454    12.620
309    13.050
416     9.405
553     9.333
        ...
487    19.440
490    12.250
267    13.590
266    10.600
243    13.750
Name: radius_mean, Length: 114, dtype: float64
```

```
plt.scatter(y_test, y_hat, c="blue", label="Datos Reales")
plt.scatter(y_test, y_hat, c="red", label="Predicciones", s=5)

plt.xlabel("Datos Reales")
plt.ylabel("Predicciones")
plt.title("Comparación entre Datos Reales y Predicciones")
plt.legend()
plt.show()
```




```
mse = mean_squared_error(y_test, y_hat)
r2 = r2_score(y_test, y_hat)

print(f"Error Cuadrático Medio: {mse:.2f}")
print(f"Coeficiente de Determinación (R^2): {r2:.2f}")
```

```
Error Cuadrático Medio: 0.00
Coeficiente de Determinación (R^2): 1.00
```

As we can see, the prediction fits perfectly to the real data. Also, we see that we do not have mean square error and therefore we have perfect R2(R2 = 1). However, we know that in real life this does not happen, since these data have been worked for educational purposes, which is why they give us such perfect values.

Isai Ambrocio - A01625101