regresionlineal2-a01633819

September 5, 2023

0.1 Actividad: Regresión Lineal 2

Andrés Alejandro Guzmán González - A01633819

Regresión lineal múltiple

Descarga la base de datos titulada "breast_cancer" disponible en canvas. Dicha base de datos contiene información sobre las características de diversos tumores.

Utiliza un modelo de regresión lineal múltiple para predecir el radio del tumor. Las variables regresoras de tu modelo deben de ser todas las variables de la base de datos.

Entrega un documento en formato PDF donde se observe la siguiente información.

1. Base de datos completa. No se observan valores faltantes. En caso de haberlos se realiza imputación simple.
2. Mostrar que las variables regresoras son independientes. En caso de no serlo realizar el procedimiento correspondiente.
3. Hipótesis nula de los coeficientes de regresión. Estadístico de prueba, distribución del estadístico de prueba. 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.
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.
5. Realiza un modelo de regresión hacia atrás (backward). Explica el criterio para ir eliminando variables del modelo.
6. Comparación entre datos reales y predicción. Análisis de los resultados.

###**Llamado a librerías:**

[140]:
```
import pandas as pd
import numpy as np
import seaborn as sns
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
import statsmodels.formula.api as smf
import statsmodels.api as sm
import matplotlib.pyplot as plt
```

```
import scipy.stats as stats
from scipy.stats import f
```

0.1.1 Paso 1.

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

Primero se inicializa él data frame, posteriormente validamos que no haya valores nulos y retiramos las variables que so se van a considerar en el modelo como es el caso de ID y Diagnóstico.

[141]:
```
df = pd.read_csv('/content/sample_data/breast_cancer.csv')
df.head()
```

[141]:

	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]
```

[142]:
```
df.isnull().sum()
```

[142]:
```
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
```

[143]:
```
df.drop(['id','diagnosis','concave points_mean','concave points_se','concave␣
 ↪points_worst'], axis = 1, inplace=True)
df.head()
```

[143]:

	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	\
0	17.99	10.38	122.80	1001.0	0.11840	
1	20.57	17.77	132.90	1326.0	0.08474	
2	19.69	21.25	130.00	1203.0	0.10960	

```
3        11.42         20.38           77.58      386.1          0.14250
4        20.29         14.34          135.10     1297.0          0.10030

   compactness_mean  concavity_mean  symmetry_mean  fractal_dimension_mean  \
0           0.27760          0.3001         0.2419                 0.07871
1           0.07864          0.0869         0.1812                 0.05667
2           0.15990          0.1974         0.2069                 0.05999
3           0.28390          0.2414         0.2597                 0.09744
4           0.13280          0.1980         0.1809                 0.05883

   radius_se  …  fractal_dimension_se  radius_worst  texture_worst  \
0     1.0950  …              0.006193         25.38          17.33
1     0.5435  …              0.003532         24.99          23.41
2     0.7456  …              0.004571         23.57          25.53
3     0.4956  …              0.009208         14.91          26.50
4     0.7572  …              0.005115         22.54          16.67

   perimeter_worst  area_worst  smoothness_worst  compactness_worst  \
0           184.60      2019.0            0.1622             0.6656
1           158.80      1956.0            0.1238             0.1866
2           152.50      1709.0            0.1444             0.4245
3            98.87       567.7            0.2098             0.8663
4           152.20      1575.0            0.1374             0.2050

   concavity_worst  symmetry_worst  fractal_dimension_worst
0           0.7119          0.4601                  0.11890
1           0.2416          0.2750                  0.08902
2           0.4504          0.3613                  0.08758
3           0.6869          0.6638                  0.17300
4           0.4000          0.2364                  0.07678

[5 rows x 27 columns]
```

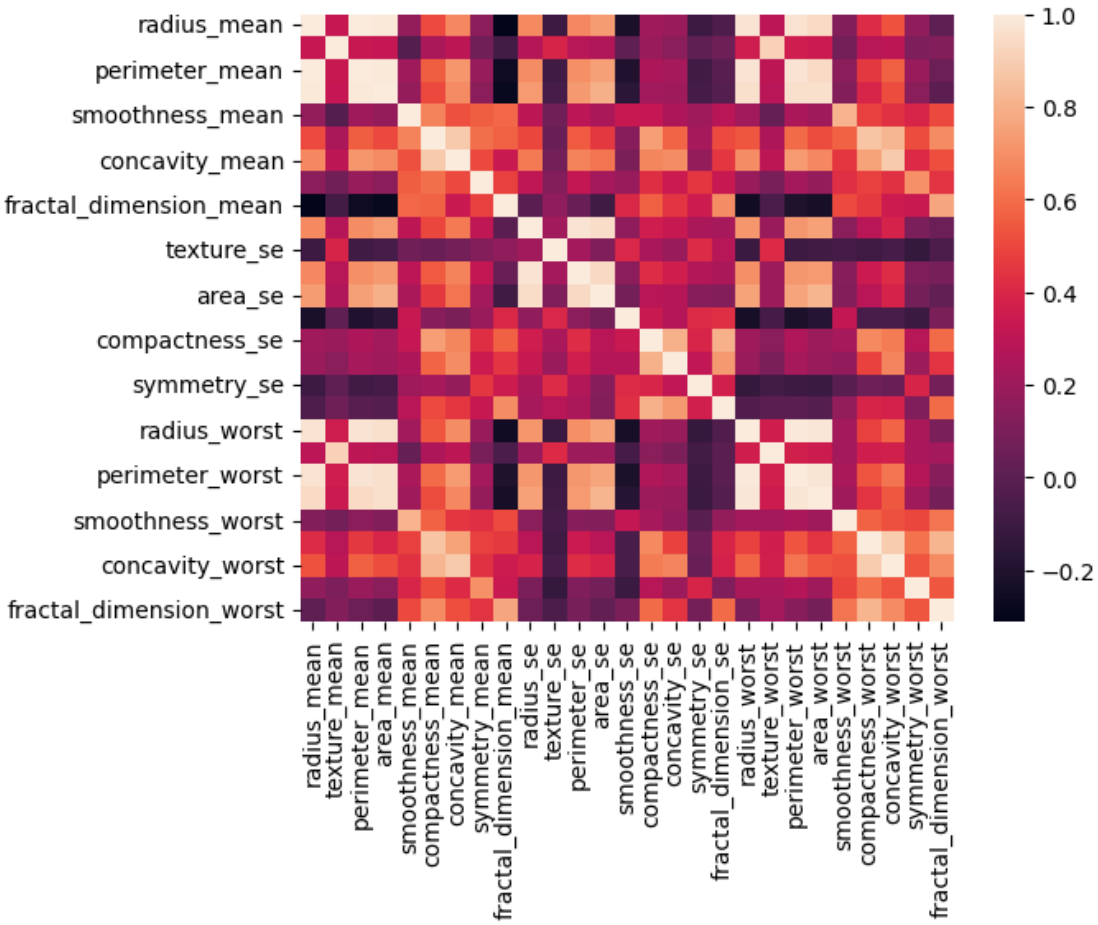
0.1.2 Paso 2.

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

Aquí se utilizó una matriz de correlación para ver de manera más visual si las variables del modelo están correlacionadas. De igual manera ubiqué en que parte del arreglo se encontraban los valores con alta correlación. Dicho esto, se procedió a estandarizar los datos.

[144]:
```
corr_matrix = df.corr()
sns.heatmap(corr_matrix, annot=False)
```

[144]: <Axes: >



[145]:
```
correlacion = df.corr()
alt_corr = np.where((correlacion > 0.95) & (correlacion < 1))
alt_corr
```

[145]:
```
(array([ 0,  0,  0,  0,  2,  2,  2,  2,  3,  3,  3,  3,  3,  9,  9, 11, 12,
        18, 18, 18, 18, 18, 20, 20, 20, 20, 20, 21, 21, 21]),
 array([ 2,  3, 18, 20,  0,  3, 18, 20,  0,  2, 18, 20, 21, 11, 12,  9,  9,
         0,  2,  3, 20, 21,  0,  2,  3, 18, 21,  3, 18, 20]))
```

[146]:
```
baja_corr = np.where((correlacion < -0.95) & (correlacion > -1))
baja_corr
```

[146]:
```
(array([], dtype=int64), array([], dtype=int64))
```

[147]:
```
scaler = StandardScaler()
```

[148]:
```
df_estandar = scaler.fit_transform(df)
```

[149]:
```
df_estandar = pd.DataFrame(df_estandar, columns=df.columns)
df_estandar
```

[149]:

	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	\
0	1.097064	-2.073335	1.269934	0.984375	1.568466	
1	1.829821	-0.353632	1.685955	1.908708	-0.826962	
2	1.579888	0.456187	1.566503	1.558884	0.942210	
3	-0.768909	0.253732	-0.592687	-0.764464	3.283553	
4	1.750297	-1.151816	1.776573	1.826229	0.280372	
..	…	…	…	…	…	
564	2.110995	0.721473	2.060786	2.343856	1.041842	
565	1.704854	2.085134	1.615931	1.723842	0.102458	
566	0.702284	2.045574	0.672676	0.577953	-0.840484	
567	1.838341	2.336457	1.982524	1.735218	1.525767	
568	-1.808401	1.221792	-1.814389	-1.347789	-3.112085	

	compactness_mean	concavity_mean	symmetry_mean	fractal_dimension_mean	\
0	3.283515	2.652874	2.217515	2.255747	
1	-0.487072	-0.023846	0.001392	-0.868652	
2	1.052926	1.363478	0.939685	-0.398008	
3	3.402909	1.915897	2.867383	4.910919	
4	0.539340	1.371011	-0.009560	-0.562450	
..	…	…	…	…	
564	0.219060	1.947285	-0.312589	-0.931027	
565	-0.017833	0.693043	-0.217664	-1.058611	
566	-0.038680	0.046588	-0.809117	-0.895587	
567	3.272144	3.296944	2.137194	1.043695	
568	-1.150752	-1.114873	-0.820070	-0.561032	

	radius_se	…	fractal_dimension_se	radius_worst	texture_worst	\
0	2.489734	…	0.907083	1.886690	-1.359293	
1	0.499255	…	-0.099444	1.805927	-0.369203	
2	1.228676	…	0.293559	1.511870	-0.023974	
3	0.326373	…	2.047511	-0.281464	0.133984	
4	1.270543	…	0.499328	1.298575	-1.466770	
..	…	…	…	…	…	
564	2.782080	…	0.167980	1.901185	0.117700	
565	1.300499	…	-0.490556	1.536720	2.047399	
566	0.184892	…	0.036727	0.561361	1.374854	
567	1.157935	…	0.904057	1.961239	2.237926	
568	-0.070279	…	-0.382754	-1.410893	0.764190	

	perimeter_worst	area_worst	smoothness_worst	compactness_worst	\
0	2.303601	2.001237	1.307686	2.616665	
1	1.535126	1.890489	-0.375612	-0.430444	
2	1.347475	1.456285	0.527407	1.082932	
3	-0.249939	-0.550021	3.394275	3.893397	

```
4           1.338539    1.220724          0.220556          -0.313395
..               …           …                 …                 …
564         1.752563    2.015301          0.378365          -0.273318
565         1.421940    1.494959         -0.691230          -0.394820
566         0.579001    0.427906         -0.809587           0.350735
567         2.303601    1.653171          1.430427           3.904848
568        -1.432735   -1.075813         -1.859019          -1.207552

     concavity_worst  symmetry_worst  fractal_dimension_worst
0           2.109526        2.750622                 1.937015
1          -0.146749       -0.243890                 0.281190
2           0.854974        1.152255                 0.201391
3           1.989588        6.046041                 4.935010
4           0.613179       -0.868353                -0.397100
..               …               …                       …
564         0.664512       -1.360158                -0.709091
565         0.236573       -0.531855                -0.973978
566         0.326767       -1.104549                -0.318409
567         3.197605        1.919083                 2.219635
568        -1.305831       -0.048138                -0.751207

[569 rows x 27 columns]
```

[150]:
```
entrenamiento, prueba = train_test_split(df_estandar, test_size=0.20,␣
 ↪random_state=42)
entrenamiento
```

[150]:
```
     radius_mean  texture_mean  perimeter_mean  area_mean  smoothness_mean  \
68     -1.447987     -0.456023       -1.366651  -1.150124         0.728714
181     1.977508      1.694187        2.089619   1.866047         1.262455
63     -1.407089     -1.263516       -1.349763  -1.120545        -1.362838
248    -0.987600      1.380033       -0.986877  -0.875668         0.014925
60     -1.123927     -1.026155       -1.129395  -0.975496         1.212639
..           …             …               …          …                …
71     -1.488033     -1.082004       -1.366651  -1.168611         0.104593
106    -0.706426     -0.223317       -0.691956  -0.689379         1.269571
270     0.046211     -0.574704       -0.068748  -0.063392        -2.282296
435    -0.041833      0.076875       -0.034972  -0.157532         0.686015
102    -0.553058      0.286311       -0.607516  -0.557982        -1.155035

     compactness_mean  concavity_mean  symmetry_mean  fractal_dimension_mean  \
68           0.700428        2.814833       1.093024                2.503828
181          3.389643        2.007548       2.129892                1.585220
63          -0.318972       -0.363081       1.932741                0.968562
248         -0.606466       -0.816190       0.311723                0.069801
60          -0.449737       -0.978777       3.400421                0.964310
..                …               …              …                       …
```

```
71          0.924055       -0.034392       0.329977                3.827870
106        -0.050051       -0.227236      -0.038768                0.340564
270        -1.470464       -1.023849      -1.108494               -1.281175
435         0.169787        0.298817      -0.520693                0.374586
102        -1.212155       -0.815688      -0.265127               -0.854476

     radius_se  …  fractal_dimension_se  radius_worst  texture_worst  \
68   -0.280696  …              2.180277     -1.234044      -0.492965
181   0.810729  …              0.567413      2.155897       1.270634
63    0.016703  …              0.766752     -1.296169      -1.049890
248  -0.561131  …             -0.444787     -0.832304       1.549097
60    0.399279  …              0.816303     -1.087016      -1.339752
..         …  …                   …            …              …
71    0.436815  …              6.859624     -1.353531      -1.629614
106  -0.357933  …              0.017058     -0.648001       0.583433
270  -0.992432  …             -0.913062     -0.281464      -0.818652
435  -0.665437  …             -0.358924      0.159621       0.834212
102  -0.767939  …             -0.855946     -0.606584       1.166414

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

     concavity_worst  symmetry_worst  fractal_dimension_worst
68          4.700669        2.147190                 1.859432
181         1.946890        1.936879                 2.463465
63         -0.635617        0.616770                 0.052877
248        -0.766109        0.822228                -0.137199
60         -1.201820        1.061659                -0.207578
..               …               …                        …
71         -0.617866       -1.046309                 1.355149
106         0.072498       -0.153294                 0.389251
270        -1.120358       -0.716282                -1.260478
435         0.646282        0.450138                 1.194443
102        -0.756514       -0.334485                -0.840426

[455 rows x 27 columns]
```

0.1.3 Paso 3.

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

[151]:
```
df.columns
```

[151]:
```
Index(['radius_mean', 'texture_mean', 'perimeter_mean', 'area_mean',
       'smoothness_mean', 'compactness_mean', 'concavity_mean',
       'symmetry_mean', 'fractal_dimension_mean', 'radius_se', 'texture_se',
       'perimeter_se', 'area_se', 'smoothness_se', 'compactness_se',
       'concavity_se', 'symmetry_se', 'fractal_dimension_se', 'radius_worst',
       'texture_worst', 'perimeter_worst', 'area_worst', 'smoothness_worst',
       'compactness_worst', 'concavity_worst', 'symmetry_worst',
       'fractal_dimension_worst'],
      dtype='object')
```

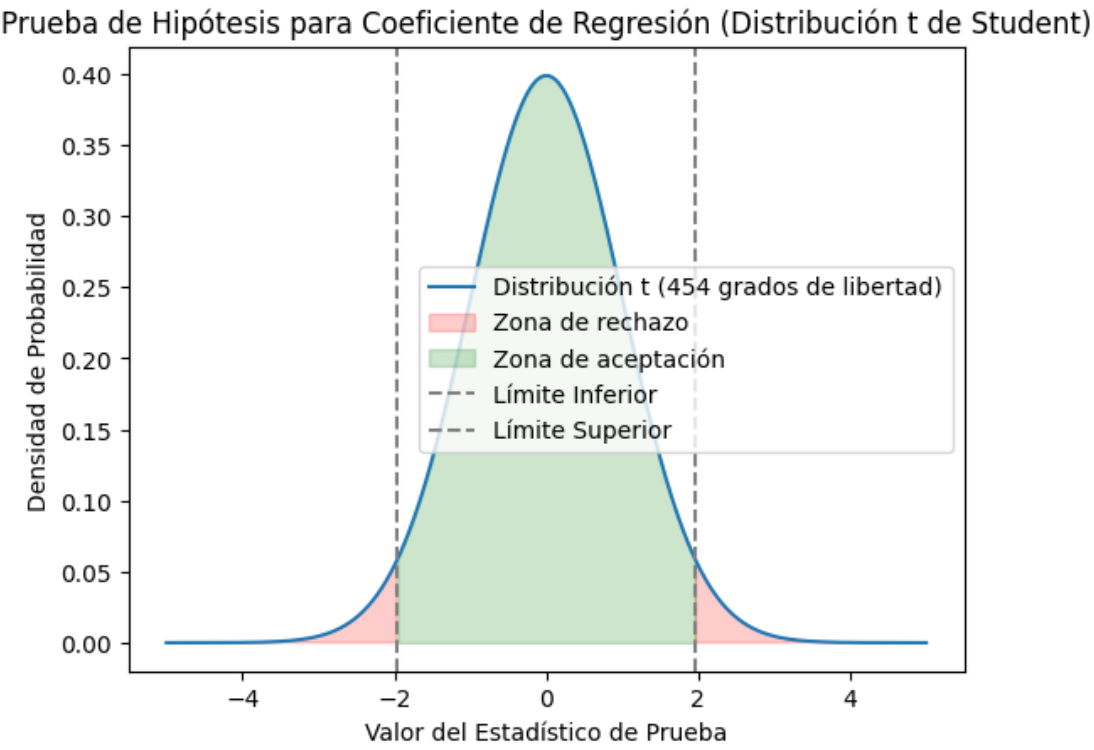
[152]:
```
modelo = smf.
 ↪ols(formula='radius_mean~texture_mean+perimeter_mean+area_mean+smoothness_mean+compactness_
 ↪data=entrenamiento)
modelo = modelo.fit()
print(modelo.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            radius_mean   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 6.611e+04
Date:                Tue, 05 Sep 2023   Prob (F-statistic):               0.00
Time:                        01:39:44   Log-Likelihood:                1240.8
No. Observations:                 455   AIC:                            -2428.
Df Residuals:                     428   BIC:                            -2316.
Df Model:                          26
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
Intercept      0.0005      0.001      0.630      0.529      -0.001
0.002
texture_mean  -0.0016      0.003     -0.598      0.550      -0.007
0.004
perimeter_mean  0.9492      0.018     54.007      0.000       0.915
0.984
area_mean      0.0715      0.013      5.299      0.000       0.045
0.098
```

```
smoothness_mean              0.0067      0.002      3.253      0.001       0.003
0.011
compactness_mean            -0.0565      0.005    -11.860      0.000      -0.066
-0.047
concavity_mean              -0.0363      0.004     -8.830      0.000      -0.044
-0.028
symmetry_mean                0.0038      0.002      2.443      0.015       0.001
0.007
fractal_dimension_mean       0.0072      0.003      2.382      0.018       0.001
0.013
radius_se                    0.0045      0.006      0.694      0.488      -0.008
0.017
texture_se                -9.373e-05     0.002     -0.058      0.953      -0.003
0.003
perimeter_se                -0.0163      0.006     -2.742      0.006      -0.028
-0.005
area_se                      0.0006      0.004      0.129      0.897      -0.008
0.009
smoothness_se                0.0014      0.001      0.958      0.338      -0.001
0.004
compactness_se              -0.0018      0.003     -0.662      0.508      -0.007
0.004
concavity_se                 0.0144      0.002      6.440      0.000       0.010
0.019
symmetry_se                  0.0044      0.002      2.462      0.014       0.001
0.008
fractal_dimension_se        -0.0032      0.002     -1.415      0.158      -0.008
0.001
radius_worst                 0.2323      0.018     12.784      0.000       0.197
0.268
texture_worst                0.0002      0.003      0.059      0.953      -0.006
0.007
perimeter_worst             -0.1139      0.015     -7.626      0.000      -0.143
-0.085
area_worst                  -0.0840      0.013     -6.369      0.000      -0.110
-0.058
smoothness_worst            -0.0049      0.002     -2.064      0.040      -0.010
-0.000
compactness_worst            0.0157      0.005      3.477      0.001       0.007
0.025
concavity_worst              0.0010      0.004      0.268      0.788      -0.007
0.009
symmetry_worst              -0.0048      0.002     -2.069      0.039      -0.009
-0.000
fractal_dimension_worst     -0.0035      0.003     -1.050      0.294      -0.010
0.003
==============================================================================
Omnibus:                       46.518   Durbin-Watson:                   2.076
```

```
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               200.782
Skew:                           0.306   Prob(JB):                     2.52e-44
Kurtosis:                       6.196   Cond. No.                         120.
==============================================================================

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

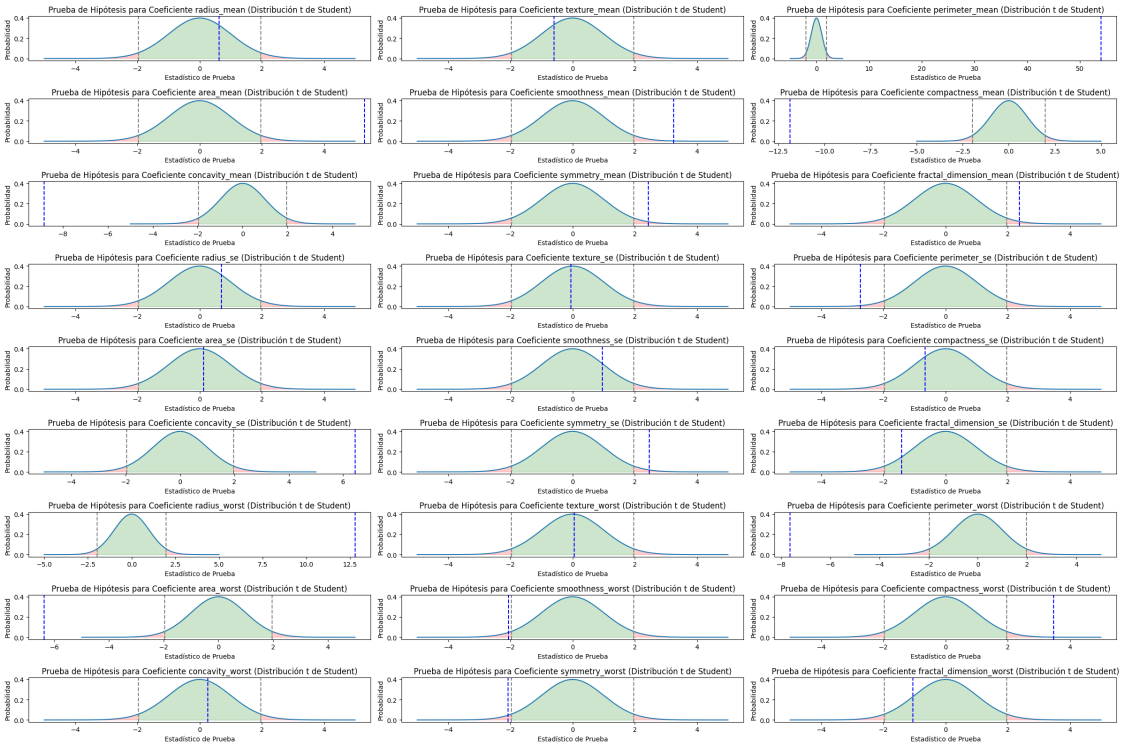
[153]:
```
# Prueba de hipótesis utilizando una distribución t de Student
confianza = 0.95
alpha = 1 - confianza
grados_libertad = len(entrenamiento) - 1
valor_critico = stats.t.ppf(1 - alpha / 2, df=grados_libertad)

# Límites de la zona de aceptación y rechazo
limite_inferior = -valor_critico
limite_superior = valor_critico

# Gráfico de la distribución t de Student
x = np.linspace(-5, 5, 1000)
pdf = stats.t.pdf(x, df=grados_libertad)
plt.plot(x, pdf, label=f'Distribución t ({grados_libertad} grados de libertad)')
plt.fill_between(x, 0, pdf, where=(x < limite_inferior) | (x >␣
 ↪limite_superior), color='red', alpha=0.2, label='Zona de rechazo')
plt.fill_between(x, 0, pdf, where=(x >= limite_inferior) & (x <=␣
 ↪limite_superior), color='green', alpha=0.2, label='Zona de aceptación')

plt.axvline(limite_inferior, color='gray', linestyle='--', label='Límite␣
 ↪Inferior')
plt.axvline(limite_superior, color='gray', linestyle='--', label='Límite␣
 ↪Superior')

plt.xlabel('Valor del Estadístico de Prueba')
plt.ylabel('Densidad de Probabilidad')
plt.title('Prueba de Hipótesis para Coeficiente de Regresión (Distribución t de␣
 ↪Student)')
plt.legend()
plt.show()
```



Aquí generé los gráficos de los valores obtenidos de todas las variables regresoras para determinar en cuáles se acepta o rechaza la hipótesis nula

[154]:
```
coef_index = 0
confianza = 0.95
alpha = 1 - confianza
grados_libertad = len(entrenamiento) - 1
valor_critico = stats.t.ppf(1 - alpha / 2, df=grados_libertad)

fig, axes = plt.subplots(nrows=9, ncols=3, figsize=(24, 16))
plt.subplots_adjust(hspace=0.5)

limite_inferior = -valor_critico
limite_superior = valor_critico

for i, fila in enumerate(axes):
    for j, ax in enumerate(fila):
        coef = modelo.params[coef_index]
        coef_std_error = modelo.bse[coef_index]
        estadistico_prueba = (coef) / coef_std_error
        x = np.linspace(-5, 5, 1000)
        pdf = stats.t.pdf(x, df=grados_libertad)
```

```
        ax.plot(x, pdf, label=f'Distribución t ({grados_libertad} grados de libertad)')
        ax.fill_between(x, 0, pdf, where=(x < limite_inferior) | (x > limite_superior), color='red', alpha=0.2, label='Zona de rechazo')
        ax.fill_between(x, 0, pdf, where=(x >= limite_inferior) & (x <= limite_superior), color='green', alpha=0.2, label='Zona de aceptación')

        ax.axvline(estadistico_prueba, color='blue', linestyle='--', label='Estadístico de Prueba')
        ax.axvline(limite_inferior, color='gray', linestyle='--', label='Límite Inferior')
        ax.axvline(limite_superior, color='gray', linestyle='--', label='Límite Superior')

        ax.set_xlabel('Estadístico de Prueba')
        ax.set_ylabel('Probabilidad')
        ax.set_title('Prueba de Hipótesis para Coeficiente '+str(df.columns[coef_index])+' (Distribución t de Student)')
        coef_index += 1
plt.tight_layout()
plt.show()
```



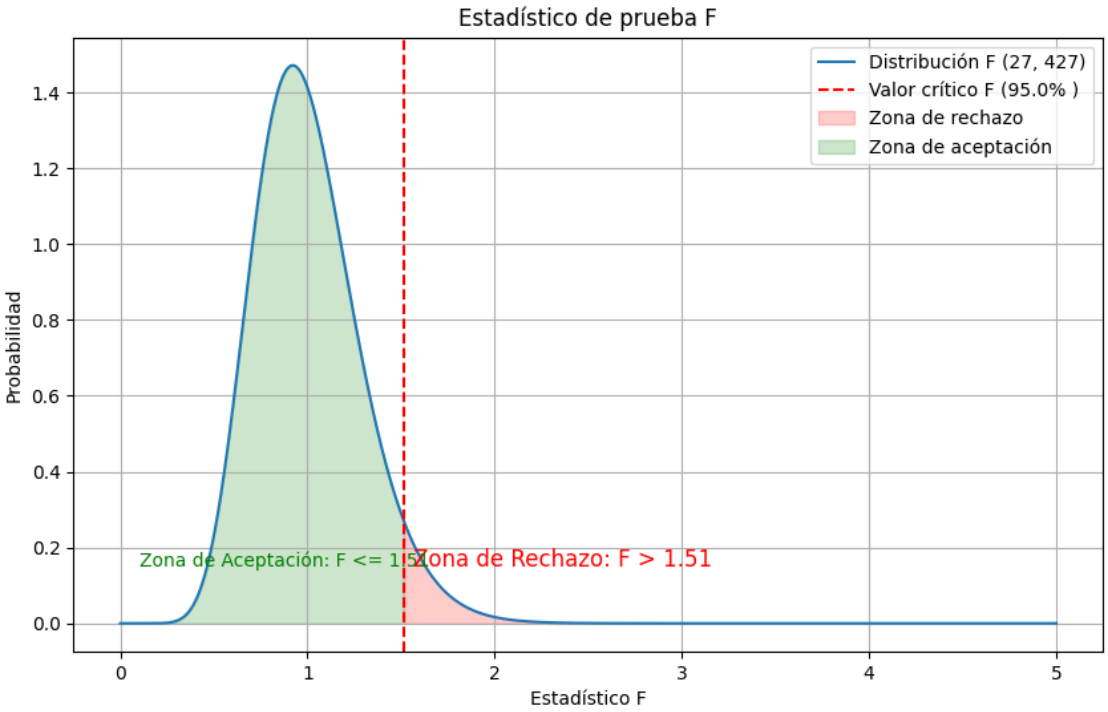
0.1.4 Paso 4.

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.

Considerando que los resultados del valor F-Static fue (6.611e+04) y que el p-valor de esta prueba dio 0, se puede determinar que hay variables regresoras que impactan o modelan el comportamiento de los datos.

A continuación se muestra la gráfica F - Fisher con el intervalo de confianza al 95%. Cabe mencionar que como el P-Valor es muy alto se encuentra en la zona de rechazo.

[155]:
```
num = 27
den = 455-(28)
confianza = 0.95
valor_critico_F = f.ppf(confianza, dfn=num,dfd=den)

rango_F = np.linspace(0, 5, 1000)
densidad_F = f.pdf(rango_F, dfn=num, dfd=den)

plt.figure(figsize=(10, 6))
plt.plot(rango_F, densidad_F, label=f'Distribución F ({num}, {den})')
plt.axvline(x=valor_critico_F, color='red', linestyle='--', label=f'Valor␣
 ↪crítico F ({confianza * 100}% )')
plt.fill_between(rango_F, densidad_F, where=((rango_F > valor_critico_F)),␣
 ↪color='red', alpha=0.2, label='Zona de rechazo')
plt.fill_between(rango_F, densidad_F, where=((rango_F <= valor_critico_F)),␣
 ↪color='green', alpha=0.2, label='Zona de aceptación')

plt.title('Estadístico de prueba F')
plt.xlabel('Estadístico F')
plt.ylabel('Probabilidad')
plt.legend()
plt.grid()

plt.text(valor_critico_F + 0.05, 0.15, f'Zona de Rechazo: F > {valor_critico_F:.
 ↪2f}', fontsize=12, color='red')
plt.text(0.1, 0.15, f'Zona de Aceptación: F <= {valor_critico_F:.2f}',␣
 ↪fontsize=10, color='green')

plt.show()
```



0.1.5 PASO 5.

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

Considerando el primer ajuste del modelo y retomando los valores del P-Valor se procede a eliminar las variables que son menos significativas en el modelo. Esto se determina discriminando las variables regresoras con el mayor P-Valor si, solo si este tiene un valor mayor a 0.05.

[156]:
```
modelo = smf.
  ↪ols(formula='radius_mean~texture_mean+perimeter_mean+area_mean+smoothness_mean+compactness_
  ↪data=entrenamiento)
modelo = modelo.fit()
print(modelo.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            radius_mean   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                6.892e+04
Date:                Tue, 05 Sep 2023   Prob (F-statistic):               0.00
Time:                        01:39:50   Log-Likelihood:                1240.8
No. Observations:                 455   AIC:                            -2430.
Df Residuals:                     429   BIC:                            -2323.
Df Model:                          25
```

```
Covariance Type:            nonrobust
==============================================================================
==========
                               coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
----------
Intercept                    0.0005      0.001      0.632      0.528      -0.001
0.002
texture_mean                -0.0015      0.002     -0.669      0.504      -0.006
0.003
perimeter_mean               0.9490      0.017     55.141      0.000       0.915
0.983
area_mean                    0.0717      0.013      5.381      0.000       0.045
0.098
smoothness_mean              0.0067      0.002      3.269      0.001       0.003
0.011
compactness_mean            -0.0565      0.005    -11.974      0.000      -0.066
-0.047
concavity_mean              -0.0363      0.004     -8.853      0.000      -0.044
-0.028
symmetry_mean                0.0038      0.002      2.512      0.012       0.001
0.007
fractal_dimension_mean       0.0072      0.003      2.388      0.017       0.001
0.013
radius_se                    0.0044      0.006      0.693      0.489      -0.008
0.017
perimeter_se                -0.0163      0.006     -2.756      0.006      -0.028
-0.005
area_se                      0.0006      0.004      0.149      0.882      -0.008
0.009
smoothness_se                0.0014      0.001      0.960      0.338      -0.001
0.004
compactness_se              -0.0018      0.003     -0.662      0.509      -0.007
0.004
concavity_se                 0.0144      0.002      6.448      0.000       0.010
0.019
symmetry_se                  0.0043      0.002      2.682      0.008       0.001
0.007
fractal_dimension_se        -0.0032      0.002     -1.417      0.157      -0.008
0.001
radius_worst                 0.2324      0.018     12.964      0.000       0.197
0.268
texture_worst             5.961e-05      0.002      0.025      0.980      -0.005
0.005
perimeter_worst             -0.1139      0.015     -7.655      0.000      -0.143
-0.085
area_worst                  -0.0842      0.013     -6.526      0.000      -0.110
```

```
-0.059
smoothness_worst           -0.0049      0.002     -2.084      0.038      -0.010
-0.000
compactness_worst           0.0157      0.005      3.482      0.001       0.007
0.025
concavity_worst             0.0011      0.004      0.271      0.786      -0.007
0.009
symmetry_worst             -0.0047      0.002     -2.198      0.028      -0.009
-0.000
fractal_dimension_worst    -0.0035      0.003     -1.053      0.293      -0.010
0.003
==============================================================================
Omnibus:                       46.499   Durbin-Watson:                   2.076
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              200.778
Skew:                           0.305   Prob(JB):                     2.52e-44
Kurtosis:                       6.196   Cond. No.                         117.
==============================================================================

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

[157]:
```
modelo = smf.
  ↪ols(formula='radius_mean~texture_mean+perimeter_mean+area_mean+smoothness_mean+compactness_
  ↪data=entrenamiento)
modelo = modelo.fit()
print(modelo.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            radius_mean   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 7.196e+04
Date:                Tue, 05 Sep 2023   Prob (F-statistic):               0.00
Time:                        01:39:50   Log-Likelihood:                1240.8
No. Observations:                 455   AIC:                            -2432.
Df Residuals:                     430   BIC:                            -2329.
Df Model:                          24
Covariance Type:            nonrobust
==============================================================================
==========
                             coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
----------
Intercept                    0.0005      0.001      0.632      0.528      -0.001
0.002
texture_mean                -0.0014      0.001     -1.674      0.095      -0.003
```

```
0.000
perimeter_mean              0.9489      0.017     55.548      0.000       0.915
0.982
area_mean                   0.0717      0.013      5.399      0.000       0.046
0.098
smoothness_mean             0.0066      0.002      3.274      0.001       0.003
0.011
compactness_mean           -0.0565      0.005    -12.026      0.000      -0.066
-0.047
concavity_mean             -0.0363      0.004     -8.913      0.000      -0.044
-0.028
symmetry_mean               0.0038      0.001      2.515      0.012       0.001
0.007
fractal_dimension_mean      0.0072      0.003      2.391      0.017       0.001
0.013
radius_se                   0.0044      0.006      0.695      0.487      -0.008
0.017
perimeter_se               -0.0163      0.006     -2.762      0.006      -0.028
-0.005
area_se                     0.0006      0.004      0.150      0.881      -0.008
0.009
smoothness_se               0.0014      0.001      0.966      0.334      -0.001
0.004
compactness_se             -0.0018      0.003     -0.662      0.508      -0.007
0.004
concavity_se                0.0144      0.002      6.471      0.000       0.010
0.019
symmetry_se                 0.0043      0.002      2.687      0.007       0.001
0.007
fractal_dimension_se       -0.0032      0.002     -1.419      0.157      -0.008
0.001
radius_worst                0.2325      0.018     13.047      0.000       0.197
0.267
perimeter_worst            -0.1139      0.015     -7.664      0.000      -0.143
-0.085
area_worst                 -0.0842      0.013     -6.553      0.000      -0.109
-0.059
smoothness_worst           -0.0049      0.002     -2.122      0.034      -0.009
-0.000
compactness_worst           0.0157      0.005      3.487      0.001       0.007
0.025
concavity_worst             0.0011      0.004      0.271      0.786      -0.007
0.009
symmetry_worst             -0.0047      0.002     -2.206      0.028      -0.009
-0.001
fractal_dimension_worst    -0.0035      0.003     -1.055      0.292      -0.010
0.003
==============================================================================
```

```
Omnibus:                       46.511   Durbin-Watson:                   2.076
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              200.888
Skew:                           0.305   Prob(JB):                     2.39e-44
Kurtosis:                       6.197   Cond. No.                         116.
==============================================================================

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

[158]:
```
modelo = smf.
  ↪ols(formula='radius_mean~texture_mean+perimeter_mean+area_mean+smoothness_mean+compactness_
  ↪data=entrenamiento)
modelo = modelo.fit()
print(modelo.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            radius_mean   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 7.525e+04
Date:                Tue, 05 Sep 2023   Prob (F-statistic):               0.00
Time:                        01:39:50   Log-Likelihood:                1240.8
No. Observations:                 455   AIC:                            -2434.
Df Residuals:                     431   BIC:                            -2335.
Df Model:                          23
Covariance Type:            nonrobust
==============================================================================
===========
                               coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
-----------
Intercept                    0.0005      0.001      0.635      0.526      -0.001
0.002
texture_mean                -0.0015      0.001     -1.692      0.091      -0.003
0.000
perimeter_mean               0.9497      0.016     58.594      0.000       0.918
0.982
area_mean                    0.0714      0.013      5.442      0.000       0.046
0.097
smoothness_mean              0.0067      0.002      3.286      0.001       0.003
0.011
compactness_mean            -0.0566      0.005    -12.126      0.000      -0.066
-0.047
concavity_mean              -0.0363      0.004     -8.922      0.000      -0.044
-0.028
symmetry_mean                0.0037      0.001      2.532      0.012       0.001
```

```
0.007
fractal_dimension_mean      0.0072      0.003      2.392      0.017       0.001
0.013
radius_se                   0.0048      0.006      0.839      0.402      -0.006
0.016
perimeter_se               -0.0161      0.006     -2.815      0.005      -0.027
-0.005
smoothness_se               0.0013      0.001      0.961      0.337      -0.001
0.004
compactness_se             -0.0018      0.003     -0.662      0.508      -0.007
0.004
concavity_se                0.0144      0.002      6.477      0.000       0.010
0.019
symmetry_se                 0.0042      0.002      2.749      0.006       0.001
0.007
fractal_dimension_se       -0.0032      0.002     -1.434      0.152      -0.008
0.001
radius_worst                0.2314      0.016     14.086      0.000       0.199
0.264
perimeter_worst            -0.1143      0.014     -7.892      0.000      -0.143
-0.086
area_worst                 -0.0832      0.011     -7.720      0.000      -0.104
-0.062
smoothness_worst           -0.0049      0.002     -2.120      0.035      -0.009
-0.000
compactness_worst           0.0157      0.004      3.510      0.000       0.007
0.025
concavity_worst             0.0011      0.004      0.271      0.787      -0.007
0.009
symmetry_worst             -0.0046      0.002     -2.243      0.025      -0.009
-0.001
fractal_dimension_worst    -0.0035      0.003     -1.047      0.296      -0.010
0.003
==============================================================================
Omnibus:                       46.310   Durbin-Watson:                   2.075
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              201.568
Skew:                           0.299   Prob(JB):                     1.70e-44
Kurtosis:                       6.206   Cond. No.                         104.
==============================================================================

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

[159]:
```
modelo = smf.
  ↪ols(formula='radius_mean~texture_mean+perimeter_mean+area_mean+smoothness_mean+compactness_
  ↪data=entrenamiento)
```

```
modelo = modelo.fit()
print(modelo.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            radius_mean   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 7.884e+04
Date:                Tue, 05 Sep 2023   Prob (F-statistic):               0.00
Time:                        01:39:50   Log-Likelihood:                1240.8
No. Observations:                 455   AIC:                            -2436.
Df Residuals:                     432   BIC:                            -2341.
Df Model:                          22
Covariance Type:            nonrobust
==============================================================================
===========
                               coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
-----------
Intercept                    0.0005      0.001      0.645      0.519      -0.001
0.002
texture_mean                -0.0015      0.001     -1.700      0.090      -0.003
0.000
perimeter_mean               0.9498      0.016     58.661      0.000       0.918
0.982
area_mean                    0.0712      0.013      5.441      0.000       0.045
0.097
smoothness_mean              0.0067      0.002      3.287      0.001       0.003
0.011
compactness_mean            -0.0569      0.005    -12.608      0.000      -0.066
-0.048
concavity_mean              -0.0356      0.003    -10.975      0.000      -0.042
-0.029
symmetry_mean                0.0037      0.001      2.532      0.012       0.001
0.007
fractal_dimension_mean       0.0071      0.003      2.379      0.018       0.001
0.013
radius_se                    0.0047      0.006      0.824      0.410      -0.007
0.016
perimeter_se                -0.0161      0.006     -2.814      0.005      -0.027
-0.005
smoothness_se                0.0013      0.001      0.940      0.348      -0.001
0.004
compactness_se              -0.0019      0.003     -0.689      0.491      -0.007
0.004
concavity_se                 0.0146      0.002      7.445      0.000       0.011
```

```
0.019
symmetry_se                0.0043      0.002      2.766      0.006       0.001
0.007
fractal_dimension_se      -0.0033      0.002     -1.532      0.126      -0.008
0.001
radius_worst               0.2315      0.016     14.111      0.000       0.199
0.264
perimeter_worst           -0.1143      0.014     -7.897      0.000      -0.143
-0.086
area_worst                -0.0832      0.011     -7.728      0.000      -0.104
-0.062
smoothness_worst          -0.0048      0.002     -2.107      0.036      -0.009
-0.000
compactness_worst          0.0163      0.004      4.055      0.000       0.008
0.024
symmetry_worst            -0.0047      0.002     -2.264      0.024      -0.009
-0.001
fractal_dimension_worst   -0.0033      0.003     -1.013      0.311      -0.010
0.003
==============================================================================
Omnibus:                       46.749   Durbin-Watson:                   2.073
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              203.636
Skew:                           0.305   Prob(JB):                    6.04e-45
Kurtosis:                       6.220   Cond. No.                        99.6
==============================================================================

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

[160]:
```
modelo = smf.
 ↪ols(formula='radius_mean~texture_mean+perimeter_mean+area_mean+smoothness_mean+compactness_
 ↪data=entrenamiento)
modelo = modelo.fit()
print(modelo.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            radius_mean   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 8.270e+04
Date:                Tue, 05 Sep 2023   Prob (F-statistic):               0.00
Time:                        01:39:51   Log-Likelihood:                1240.5
No. Observations:                 455   AIC:                            -2437.
Df Residuals:                     433   BIC:                            -2346.
Df Model:                          21
Covariance Type:            nonrobust
==============================================================================
```

```
==========
                            coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
----------
Intercept                  0.0005      0.001      0.641      0.522      -0.001
0.002
texture_mean              -0.0015      0.001     -1.695      0.091      -0.003
0.000
perimeter_mean             0.9494      0.016     58.709      0.000       0.918
0.981
area_mean                  0.0715      0.013      5.476      0.000       0.046
0.097
smoothness_mean            0.0067      0.002      3.335      0.001       0.003
0.011
compactness_mean          -0.0576      0.004    -13.157      0.000      -0.066
-0.049
concavity_mean            -0.0351      0.003    -11.141      0.000      -0.041
-0.029
symmetry_mean              0.0037      0.001      2.519      0.012       0.001
0.007
fractal_dimension_mean     0.0073      0.003      2.472      0.014       0.002
0.013
radius_se                  0.0052      0.006      0.929      0.354      -0.006
0.016
perimeter_se              -0.0167      0.006     -2.962      0.003      -0.028
-0.006
smoothness_se              0.0010      0.001      0.772      0.441      -0.002
0.004
concavity_se               0.0141      0.002      7.854      0.000       0.011
0.018
symmetry_se                0.0041      0.002      2.695      0.007       0.001
0.007
fractal_dimension_se      -0.0041      0.002     -2.212      0.027      -0.008
-0.000
radius_worst               0.2313      0.016     14.108      0.000       0.199
0.264
perimeter_worst           -0.1131      0.014     -7.876      0.000      -0.141
-0.085
area_worst                -0.0838      0.011     -7.830      0.000      -0.105
-0.063
smoothness_worst          -0.0046      0.002     -2.032      0.043      -0.009
-0.000
compactness_worst          0.0149      0.003      4.302      0.000       0.008
0.022
symmetry_worst            -0.0045      0.002     -2.218      0.027      -0.009
-0.001
fractal_dimension_worst   -0.0028      0.003     -0.876      0.382      -0.009
```

```
0.003
==============================================================================
Omnibus:                       47.273   Durbin-Watson:                   2.072
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              205.057
Skew:                           0.315   Prob(JB):                     2.97e-45
Kurtosis:                       6.228   Cond. No.                         97.2
==============================================================================

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

[161]:
```
modelo = smf.
  ↪ols(formula='radius_mean~texture_mean+perimeter_mean+area_mean+smoothness_mean+compactness_
  ↪data=entrenamiento)
modelo = modelo.fit()
print(modelo.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            radius_mean   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 8.692e+04
Date:                Tue, 05 Sep 2023   Prob (F-statistic):               0.00
Time:                        01:39:51   Log-Likelihood:                1240.2
No. Observations:                 455   AIC:                            -2438.
Df Residuals:                     434   BIC:                            -2352.
Df Model:                          20
Covariance Type:            nonrobust
===============================================================================
===========
                              coef    std err          t      P>|t|      [0.025
0.975]
-------------------------------------------------------------------------------
-----------
Intercept                   0.0005      0.001      0.642      0.521      -0.001
0.002
texture_mean               -0.0014      0.001     -1.631      0.104      -0.003
0.000
perimeter_mean              0.9490      0.016     58.741      0.000       0.917
0.981
area_mean                   0.0723      0.013      5.549      0.000       0.047
0.098
smoothness_mean             0.0063      0.002      3.250      0.001       0.002
0.010
compactness_mean           -0.0573      0.004    -13.145      0.000      -0.066
-0.049
concavity_mean             -0.0352      0.003    -11.198      0.000      -0.041
```

```
-0.029
symmetry_mean                0.0039      0.001      2.671      0.008       0.001
0.007
fractal_dimension_mean       0.0073      0.003      2.471      0.014       0.001
0.013
radius_se                    0.0055      0.006      0.983      0.326      -0.006
0.017
perimeter_se                -0.0169      0.006     -3.017      0.003      -0.028
-0.006
concavity_se                 0.0141      0.002      7.873      0.000       0.011
0.018
symmetry_se                  0.0046      0.001      3.276      0.001       0.002
0.007
fractal_dimension_se        -0.0038      0.002     -2.094      0.037      -0.007
-0.000
radius_worst                 0.2303      0.016     14.099      0.000       0.198
0.262
perimeter_worst             -0.1126      0.014     -7.854      0.000      -0.141
-0.084
area_worst                  -0.0838      0.011     -7.833      0.000      -0.105
-0.063
smoothness_worst            -0.0035      0.002     -1.969      0.050      -0.007
-6.47e-06
compactness_worst            0.0147      0.003      4.270      0.000       0.008
0.022
symmetry_worst              -0.0051      0.002     -2.657      0.008      -0.009
-0.001
fractal_dimension_worst     -0.0031      0.003     -0.989      0.323      -0.009
0.003
==============================================================================
Omnibus:                       48.171   Durbin-Watson:                   2.076
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              214.633
Skew:                           0.315   Prob(JB):                     2.47e-47
Kurtosis:                       6.305   Cond. No.                         97.2
==============================================================================

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

[162]:
```
modelo = smf.
  ↪ols(formula='radius_mean~texture_mean+perimeter_mean+area_mean+smoothness_mean+compactness_
  ↪data=entrenamiento)
modelo = modelo.fit()
print(modelo.summary())
```

```
                            OLS Regression Results
==============================================================================
```

Dep. Variable:	radius_mean	R-squared:	1.000
Model:	OLS	Adj. R-squared:	1.000
Method:	Least Squares	F-statistic:	9.150e+04
Date:	Tue, 05 Sep 2023	Prob (F-statistic):	0.00
Time:	01:39:51	Log-Likelihood:	1239.7
No. Observations:	455	AIC:	-2439.
Df Residuals:	435	BIC:	-2357.
Df Model:	19		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.0005	0.001	0.630	0.529	-0.001	0.002
texture_mean	-0.0014	0.001	-1.656	0.098	-0.003	0.000
perimeter_mean	0.9470	0.016	59.086	0.000	0.915	0.978
area_mean	0.0729	0.013	5.604	0.000	0.047	0.098
smoothness_mean	0.0066	0.002	3.440	0.001	0.003	0.010
compactness_mean	-0.0575	0.004	-13.201	0.000	-0.066	-0.049
concavity_mean	-0.0345	0.003	-11.274	0.000	-0.041	-0.028
symmetry_mean	0.0040	0.001	2.794	0.005	0.001	0.007
fractal_dimension_mean	0.0070	0.003	2.389	0.017	0.001	0.013
perimeter_se	-0.0117	0.002	-6.531	0.000	-0.015	-0.008
concavity_se	0.0138	0.002	7.818	0.000	0.010	0.017
symmetry_se	0.0046	0.001	3.293	0.001	0.002	0.007
fractal_dimension_se	-0.0033	0.002	-1.896	0.059	-0.007	0.000
radius_worst	0.2401	0.013	18.573	0.000	0.215	0.266
perimeter_worst	-0.1212	0.011	-10.673	0.000	-0.143	-0.099
area_worst	-0.0840	0.011	-7.848	0.000	-0.105	-0.063
smoothness_worst	-0.0036	0.002	-1.975	0.049	-0.007	

```
-1.64e-05
compactness_worst           0.0150      0.003      4.341      0.000       0.008
0.022
symmetry_worst             -0.0052      0.002     -2.703      0.007      -0.009
-0.001
fractal_dimension_worst    -0.0034      0.003     -1.098      0.273      -0.010
0.003
==============================================================================
Omnibus:                       48.888   Durbin-Watson:                   2.085
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              229.479
Skew:                           0.301   Prob(JB):                     1.48e-50
Kurtosis:                       6.427   Cond. No.                         92.2
==============================================================================

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

[163]:
```
modelo = smf.
  ↪ols(formula='radius_mean~texture_mean+perimeter_mean+area_mean+smoothness_mean+compactness_
  ↪data=entrenamiento)
modelo = modelo.fit()
print(modelo.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            radius_mean   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 9.653e+04
Date:                Tue, 05 Sep 2023   Prob (F-statistic):               0.00
Time:                        01:39:51   Log-Likelihood:                 1239.1
No. Observations:                 455   AIC:                            -2440.
Df Residuals:                     436   BIC:                            -2362.
Df Model:                          18
Covariance Type:            nonrobust
==============================================================================
==========
                             coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
----------
Intercept                  0.0005      0.001      0.616      0.538      -0.001
0.002
texture_mean              -0.0014      0.001     -1.659      0.098      -0.003
0.000
perimeter_mean             0.9471      0.016     59.078      0.000       0.916
0.979
area_mean                  0.0726      0.013      5.581      0.000       0.047
```

```
0.098
smoothness_mean             0.0067      0.002      3.509      0.000       0.003
0.010
compactness_mean           -0.0557      0.004    -13.865      0.000      -0.064
-0.048
concavity_mean             -0.0350      0.003    -11.567      0.000      -0.041
-0.029
symmetry_mean               0.0041      0.001      2.862      0.004       0.001
0.007
fractal_dimension_mean      0.0050      0.002      2.181      0.030       0.000
0.010
perimeter_se               -0.0117      0.002     -6.551      0.000      -0.015
-0.008
concavity_se                0.0144      0.002      8.465      0.000       0.011
0.018
symmetry_se                 0.0051      0.001      3.785      0.000       0.002
0.008
fractal_dimension_se       -0.0042      0.002     -2.684      0.008      -0.007
-0.001
radius_worst                0.2384      0.013     18.573      0.000       0.213
0.264
perimeter_worst            -0.1204      0.011    -10.624      0.000      -0.143
-0.098
area_worst                 -0.0830      0.011     -7.780      0.000      -0.104
-0.062
smoothness_worst           -0.0039      0.002     -2.172      0.030      -0.007
-0.000
compactness_worst           0.0124      0.003      4.919      0.000       0.007
0.017
symmetry_worst             -0.0056      0.002     -2.988      0.003      -0.009
-0.002
==============================================================================
Omnibus:                       49.686   Durbin-Watson:                   2.072
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              247.933
Skew:                           0.282   Prob(JB):                     1.45e-54
Kurtosis:                       6.572   Cond. No.                         90.6
==============================================================================

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

[164]:
```
modelo = smf.
  ↪ols(formula='radius_mean~perimeter_mean+area_mean+smoothness_mean+compactness_mean+concavit
  ↪data=entrenamiento)
modelo = modelo.fit()
print(modelo.summary())
```

OLS Regression Results

Dep. Variable:	radius_mean	R-squared:	1.000
Model:	OLS	Adj. R-squared:	1.000
Method:	Least Squares	F-statistic:	1.018e+05
Date:	Tue, 05 Sep 2023	Prob (F-statistic):	0.00
Time:	01:39:51	Log-Likelihood:	1237.7
No. Observations:	455	AIC:	-2439.
Df Residuals:	437	BIC:	-2365.
Df Model:	17		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.0005	0.001	0.656	0.512	-0.001	0.002
perimeter_mean	0.9482	0.016	59.078	0.000	0.917	0.980
area_mean	0.0722	0.013	5.542	0.000	0.047	0.098
smoothness_mean	0.0072	0.002	3.791	0.000	0.003	0.011
compactness_mean	-0.0556	0.004	-13.832	0.000	-0.064	-0.048
concavity_mean	-0.0353	0.003	-11.660	0.000	-0.041	-0.029
symmetry_mean	0.0039	0.001	2.704	0.007	0.001	0.007
fractal_dimension_mean	0.0053	0.002	2.308	0.021	0.001	0.010
perimeter_se	-0.0118	0.002	-6.584	0.000	-0.015	-0.008
concavity_se	0.0144	0.002	8.503	0.000	0.011	0.018
symmetry_se	0.0048	0.001	3.633	0.000	0.002	0.007
fractal_dimension_se	-0.0042	0.002	-2.684	0.008	-0.007	-0.001
radius_worst	0.2364	0.013	18.462	0.000	0.211	0.262
perimeter_worst	-0.1201	0.011	-10.576	0.000	-0.142	-0.098
area_worst	-0.0820	0.011	-7.687	0.000	-0.103	-0.061
smoothness_worst	-0.0042	0.002	-2.363	0.019	-0.008	

```
-0.001
compactness_worst          0.0119      0.003      4.746      0.000       0.007
0.017
symmetry_worst            -0.0052      0.002     -2.785      0.006      -0.009
-0.002
==============================================================================
Omnibus:                       51.478   Durbin-Watson:                   2.099
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              276.926
Skew:                           0.270   Prob(JB):                     7.35e-61
Kurtosis:                       6.784   Cond. No.                         89.9
==============================================================================

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

0.1.6 Paso 6.

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

Teniendo el ajuste final procedí a crear la función que represente al modelo y posteriormente se hace el análisis gráfico de los errores. Esto se hace calculando la diferencia de las predicciones y los valores reales.

Posteriormente, grafiqué el valor real vs. el valor predecido. En este caso se espera que los datos se ajusten lo mejor posible a la recta $y = x$.

También se graficaron las diferencias de los errores con el fin de ver que tan dispersos se encuentran con respecto de 0.

Además, se realizó el histograma de residuos y un QQ plot.

Todos estos gráficos me permitieron validar que el modelo tiene un excelente ajuste. Pues al comparar las predicciones con los valores reales, estos se mantienen muy cercanos a la línea de tendencia. Así mismo, esto se confirma en la dispersión de errores, pues estos se acotan entre 0.05 y -0.150

[165]:

```
y_pred = modelo.params[1]*prueba['perimeter_mean'] + modelo.
 ↪params[2]*prueba['area_mean'] + modelo.params[3]*prueba['smoothness_mean'] +␣
 ↪modelo.params[4]*prueba['compactness_mean'] + modelo.
 ↪params[5]*prueba['concavity_mean'] + modelo.
 ↪params[6]*prueba['symmetry_mean'] + modelo.params[7]*prueba['symmetry_mean']␣
 ↪+ modelo.params[8]*prueba['perimeter_se'] + modelo.
 ↪params[9]*prueba['concavity_se'] + modelo.params[10]*prueba['symmetry_se'] +␣
 ↪modelo.params[11]*prueba['fractal_dimension_se'] + modelo.
 ↪params[12]*prueba['radius_worst'] + modelo.
 ↪params[13]*prueba['perimeter_worst'] + modelo.
 ↪params[14]*prueba['area_worst'] + modelo.
 ↪params[15]*prueba['smoothness_worst'] + modelo.
 ↪params[16]*prueba['compactness_worst'] + modelo.
 ↪params[17]*prueba['symmetry_worst']
y_pred
```

[165]:
```
204   -0.452369
70     1.367122
131    0.399405
431   -0.507333
540   -0.742339
         …
486    0.152368
75     0.550460
249   -0.737218
238    0.007633
265    1.839478
Length: 114, dtype: float64
```

[166]:
```
tabla=pd.DataFrame({'Real': prueba['radius_mean'], 'Prediccion': y_pred,␣
 ↪'Errores': prueba['radius_mean']-y_pred})
tabla
```

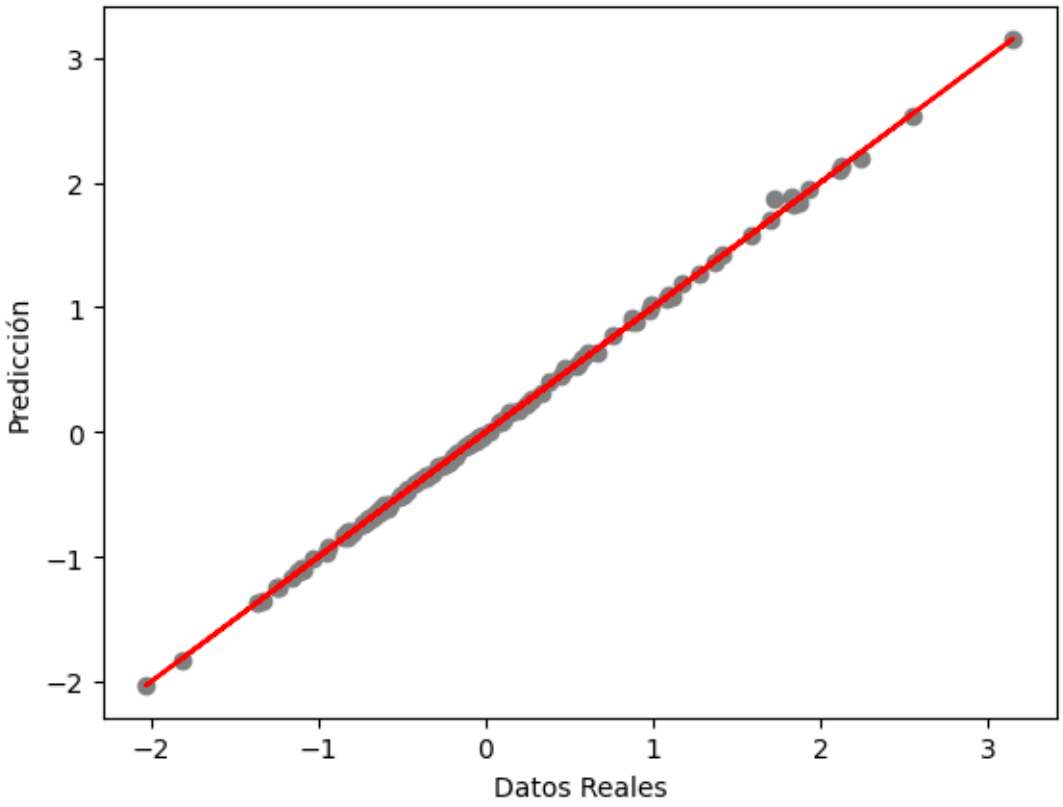
[166]:

	Real	Prediccion	Errores
204	-0.470694	-0.452369	-0.018326
70	1.366877	1.367122	-0.000244
131	0.378508	0.399405	-0.020897
431	-0.490575	-0.507333	0.016758
540	-0.734828	-0.742339	0.007512
..	…	…	…
486	0.145616	0.152368	-0.006751
75	0.551757	0.550460	0.001296
249	-0.740508	-0.737218	-0.003290
238	0.026330	0.007633	0.018697
265	1.875263	1.839478	0.035785

[114 rows x 3 columns]

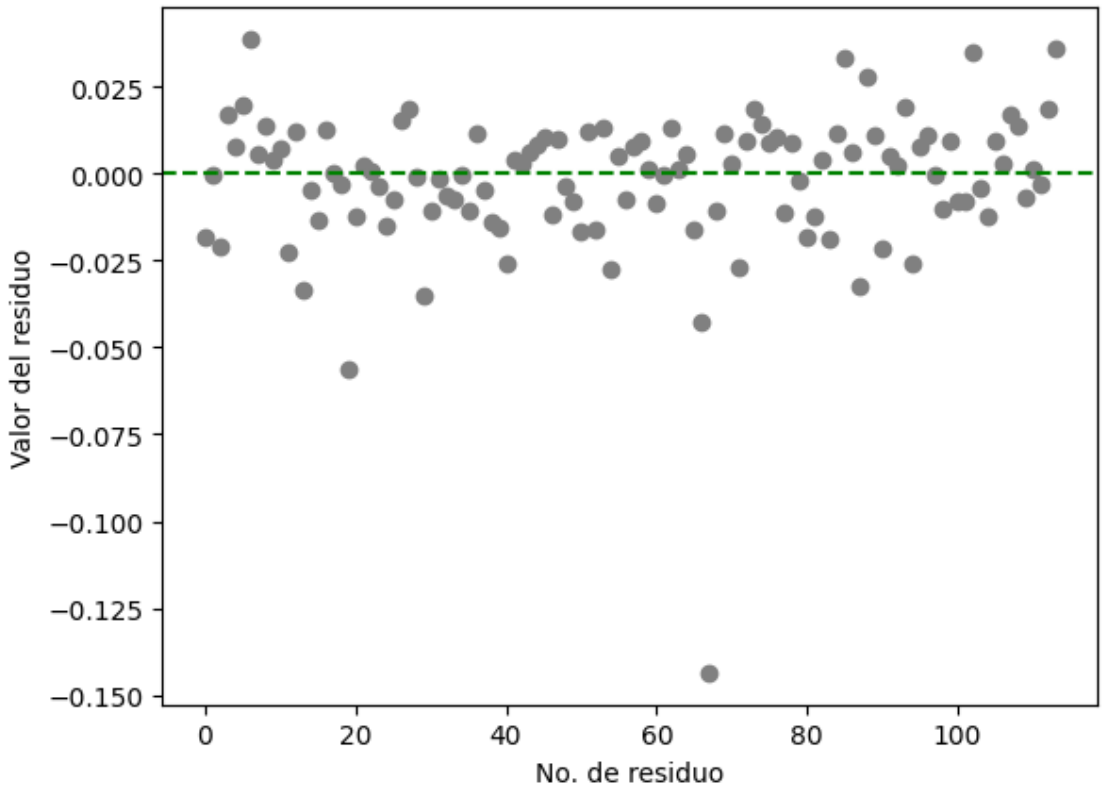
[167]:
```
plt.scatter(prueba['radius_mean'], y_pred, color='gray')
plt.plot(prueba['radius_mean'],prueba['radius_mean'], color='red')
plt.xlabel("Datos Reales")
plt.ylabel("Predicción")
```

[167]: Text(0, 0.5, 'Predicción')



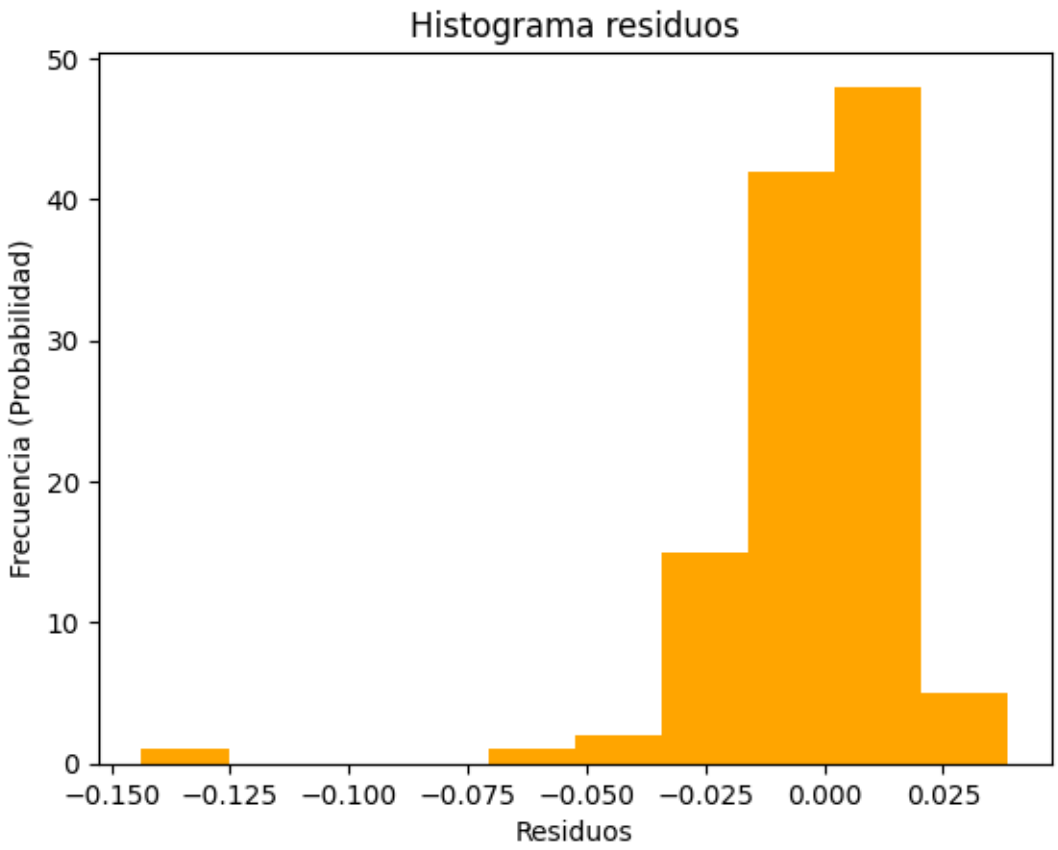
[168]:
```
plt.scatter(range(tabla.shape[0]),tabla['Errores'], color='gray')
plt.axhline(y=0, linestyle='--', color='green')
plt.xlabel("No. de residuo")
plt.ylabel("Valor del residuo")
```

[168]: Text(0, 0.5, 'Valor del residuo')

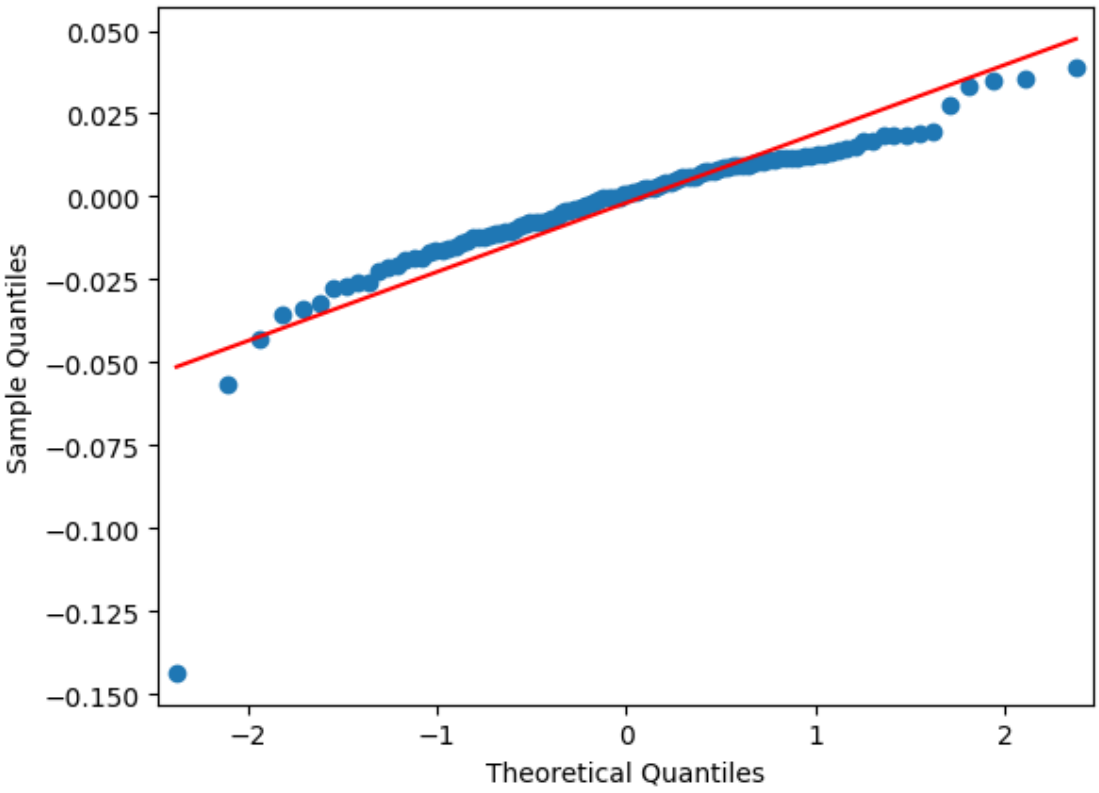


[169]:
```
plt.hist(x=tabla['Errores'], color='orange')
plt.title('Histograma residuos')
plt.xlabel("Residuos")
plt.ylabel("Frecuencia (Probabilidad)")
```

[169]: Text(0, 0.5, 'Frecuencia (Probabilidad)')



[170]:
```
QQ = sm.qqplot(tabla['Errores'], stats.norm, line='s')
```


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