Linear Regression for Soccer squad data, to predict goals
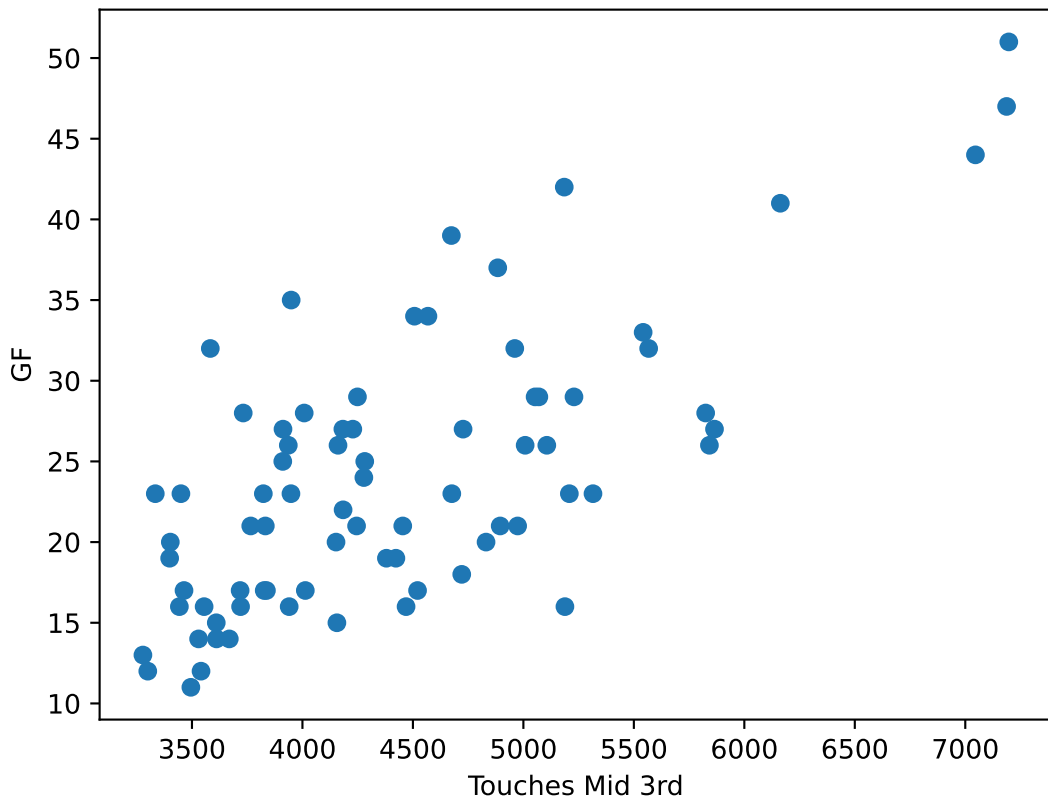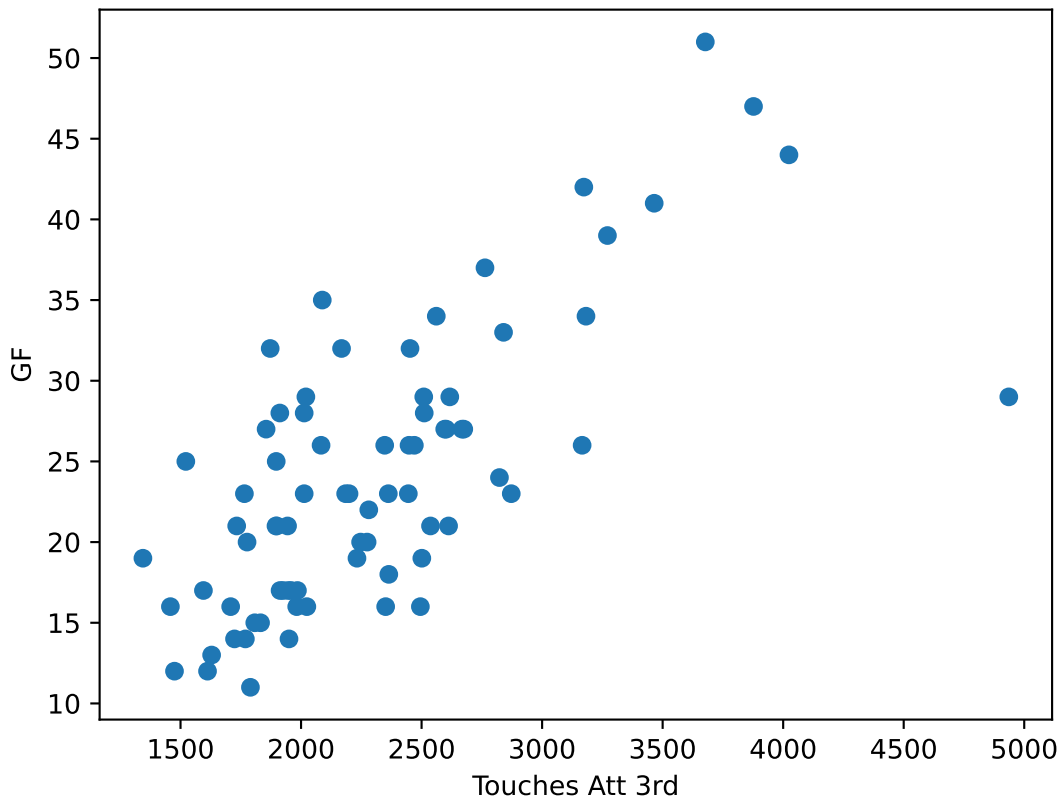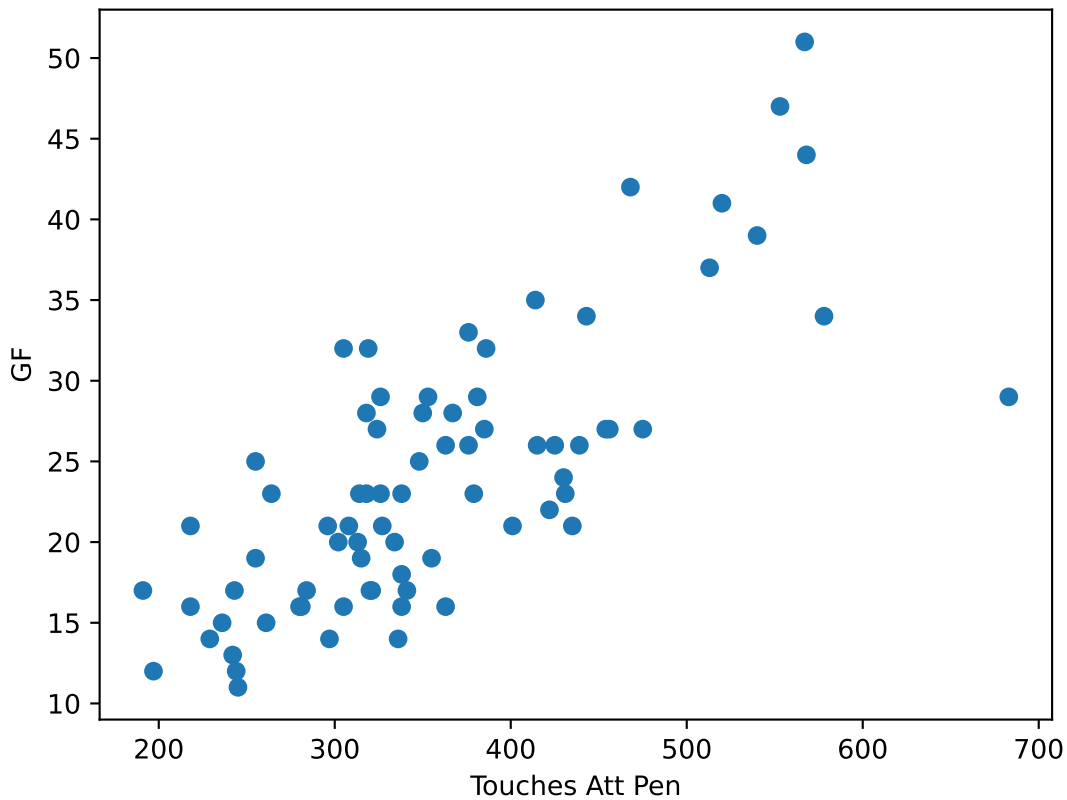
Touches Mid 3rd vs GF

Touches Att 3rd vs GF

Reduced Final Model:
1. Touches Mid 3rd
2. Touches Att 3rd
3. Touches Att PenModel Summary:

```
                    OLS Regression Results
==============================================================================
Dep. Variable:              GF   R-squared:                       0.690
Model:                     OLS   Adj. R-squared:                  0.677
Method:          Least Squares   F-statistic:                     53.35
Date:         Thu, 02 Jan 2025   Prob (F-statistic):           2.88e-18
Time:                 12:28:48   Log-Likelihood:                 -225.48
No. Observations:           76   AIC:                             459.0
Df Residuals:               72   BIC:                             468.3
Df Model:                    3
Covariance Type:     nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          -9.9428      2.965     -3.354      0.001     -15.852      -4.033
Touches Mid 3rd  0.0052      0.001      5.072      0.000       0.003       0.007
Touches Att 3rd -0.0069      0.003     -2.566      0.012      -0.012      -0.002
Touches Att Pen  0.0755      0.014      5.300      0.000       0.047       0.104
==============================================================================
Omnibus:                     0.931   Durbin-Watson:                   1.984
Prob(Omnibus):               0.628   Jarque-Bera (JB):                0.901
Skew:                        0.254   Prob(JB):                        0.637
Kurtosis:                    2.834   Cond. No.                     2.74e+04
==============================================================================
```

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.74e+04. This might indicate that there are
strong multicollinearity or other numerical problems.VIF for Reduced Model:
      Variable   VIF
Touches Mid 3rd 2.670479
Touches Att 3rd 9.307427
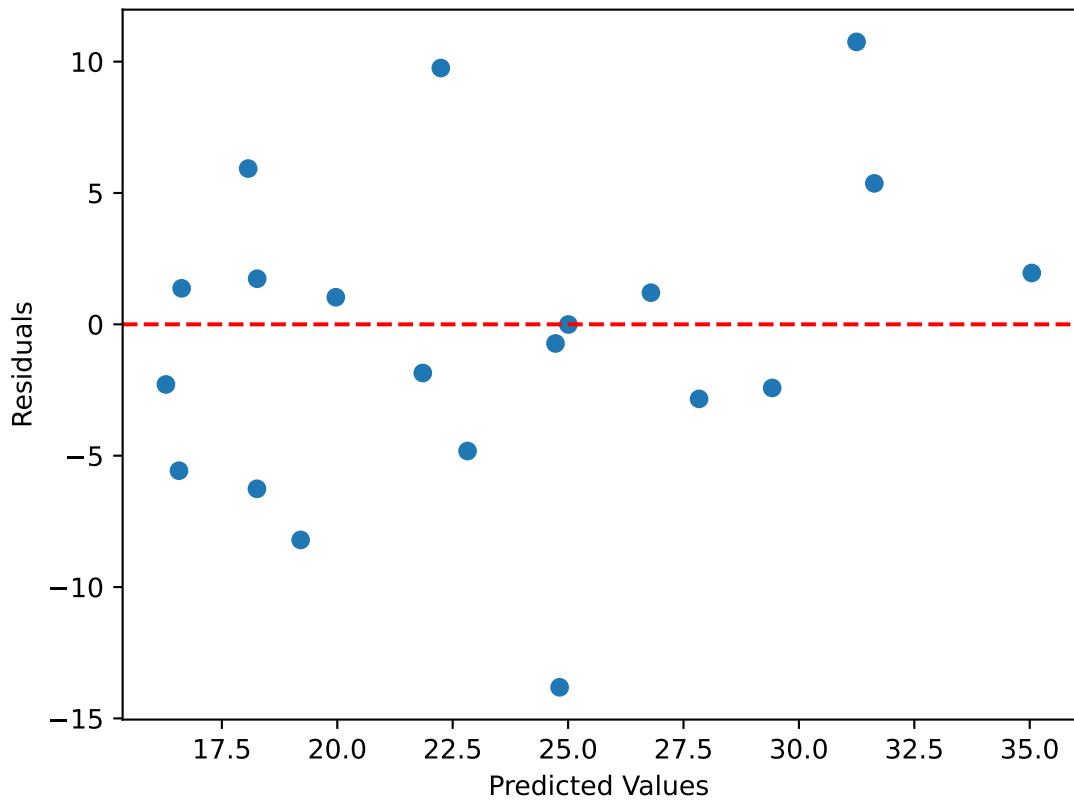Touches Att Pen 6.356137Mean Squared Error on Test Set: 33.04
Breusch-Pagan test (Equality of variance): p-value: 0.674
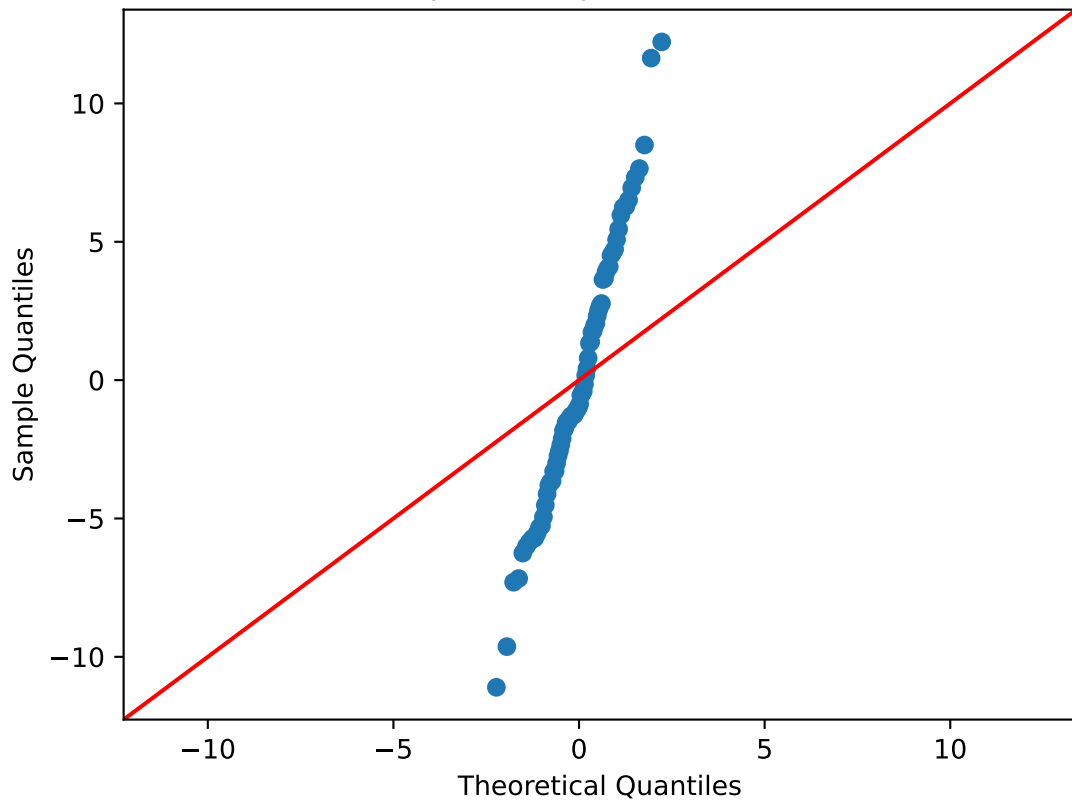Equality of Variance holds (homoscedasticity).

Confidence Intervals for Betas

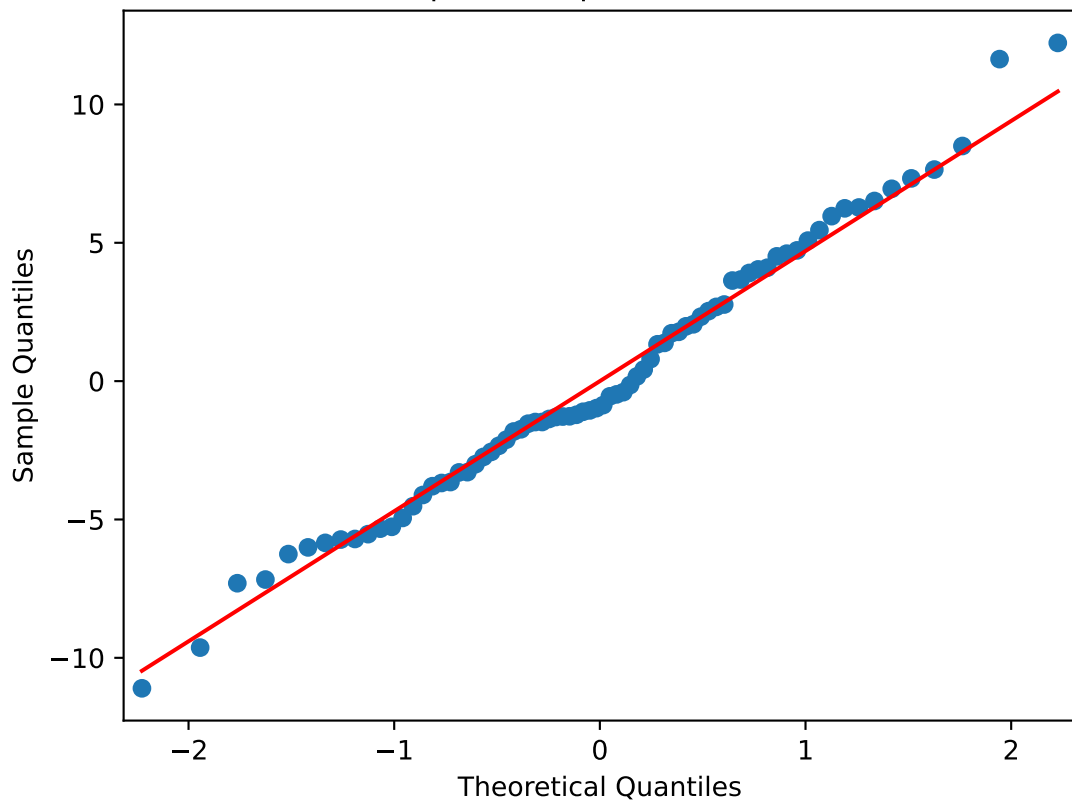| Lower CI	Upper CI
-15.852475896057879	-4.03316697479793
0.00313047097442479	0.0071842932198524015
-0.012187383876821691	-0.0015313694849647883
0.047095854671667735	0.10388023040202193

QQ Plot with standardized line
Shapiro-Wilk p-value: 0.8969

Predictions vs Actuals

Predicted	Actual
21.852409994668502	20.0
22.822161426543115	18.0
35.044896227676816	37.0
16.571375662929693	11.0
31.633708982523185	37.0
18.264009520077668	20.0
16.287798552612113	14.0
29.42080356834676	27.0
22.242699346984605	32.0
26.79417471628221	28.0
18.260023340705686	12.0
19.968315833363885	21.0
31.248705006453225	42.0
24.81638000201269	11.0
19.20646533718163	11.0
25.005952784363043	25.0
18.070093522622017	24.0
27.837771196253026	25.0
16.628456904679226	18.0
24.7270379539621	24.0