6 )

reg\_model = lm(EQI ~ SDP)

> summary(reg\_model)

Call:

lm(formula = EQI ~ SDP)

Residuals:

Min 1Q Median 3Q Max

-6.372 -2.012 -1.328 -0.517 146.074

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Coefficients | Estimate | Std. Error | T values | Pr(>|t|) |
| Intercept | 2.047 | 0.1922 | 10.651 | < 2e-16 |
| SDP | 2.851e-06 | 3.396e-07 | 8.398 | < 2e-16 |

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.647 on 2767 degrees of freedom

Multiple R-squared: 0.02486, Adjusted R-squared: 0.0245

F-statistic: 70.53 on 1 and 2767 DF, p-value: < 2.2e-16

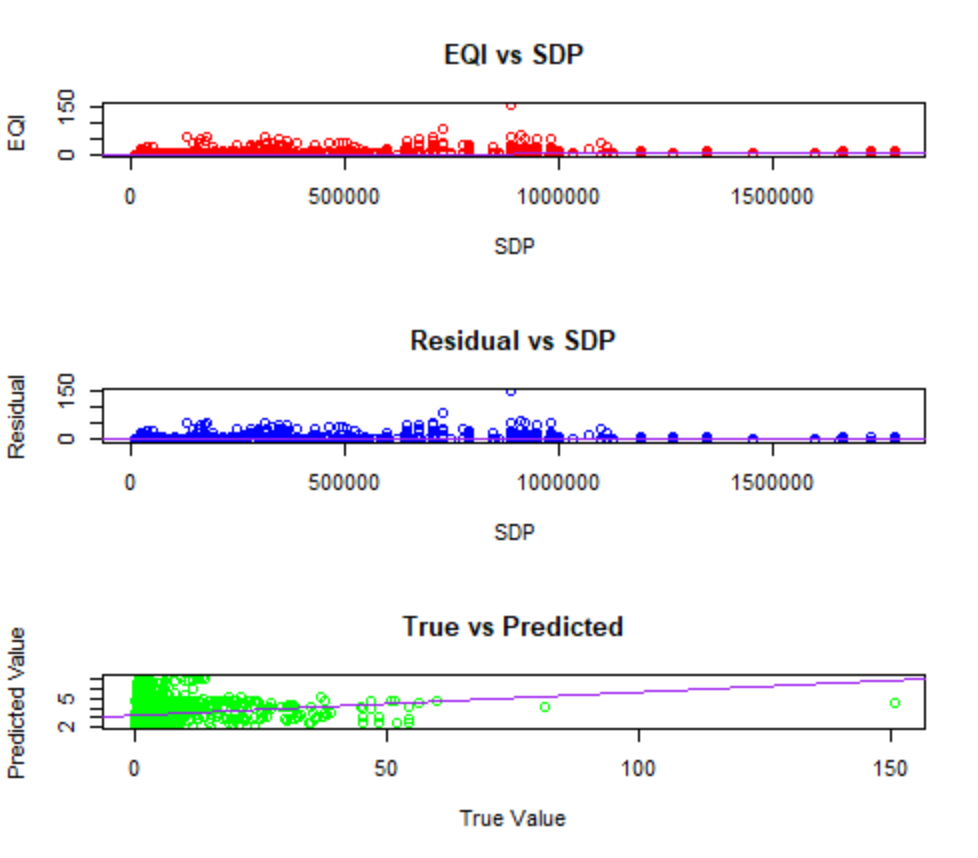
Interpretation In English:

7) The three plots are related to each other in that they all provide information about the model's goodness of fit. If the model is a good fit, we expect to see a random scatter of residuals around the fitted line in the first and second plots and a tight cluster of points around the diagonal line in the third plot.

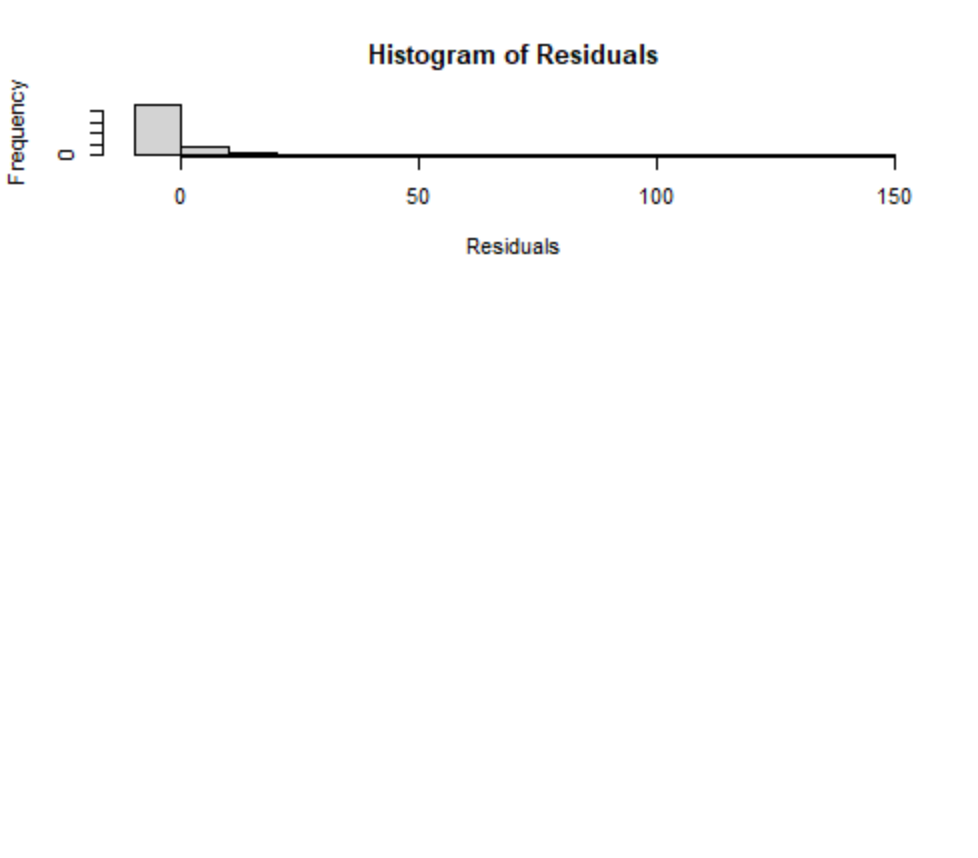
The first scatter plot shows the overall relationship between the dependent and independent variables and gives an idea of the direction and strength of the relationship. Here the red points are scattered randomly around the purple line.

The second scatter plot shows how well the model captures the variation in the dependent variable that is not explained by the independent variable, i.e., the variation of residual with changing independent variable values. Two regression model assumptions regarding the expected value and variance of residual can be verified from the second plot.

The third scatter plot shows how well the model predicts the actual values of the dependent variable. Since a large cluster of scatter points lie on the left side of the plot, it suggests there is very little variation in expected and obtained values. If there is a systematic pattern in the residuals or a wide scatter of points around the diagonal line, it suggests that the model may not be a good fit and may need to be revised.



8)





It can be verified that the sum of residuals is 0.

9) Here, SDPx is SDP ^ x, and GINI is the Gini index

Call:

lm(formula = EQI ~ SDP1 + SDP2 + SDP3 + GINI)

Residuals:

Min 1Q Median 3Q Max

-8.209 -2.433 -1.043 0.283 145.659

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Coefficients | Estimate | Std. Errors | T value | Pr(>|t|) |
| Intercept | -3.887 | 0.7367 | -5.276 | 1.44e-07 |
| SDP1 | 7.128e-06 | 2.465e-06 | 2.892 | 0.00387 |
| SDP2 | 6.417e-13 | 3.850e-12 | 0.167 | 0.86762 |
| SDP3 | -2.531e-18 | 1.572e-18 | -1.610 | 0.10743 |
| GINI | 16.60 | 2.225 | 7.461 | 1.19e-13 |

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.576 on 2447 degrees of freedom

(317 observations deleted due to missingness)

Multiple R-squared: 0.07145, Adjusted R-squared: 0.06993

F-statistic: 47.07 on 4 and 2447 DF, p-value: < 2.2e-16