

NAME	REGISTRATION NO.
George Ouma	KPSK-TUM-08
Asha Kahindi	KPSK-TUM-07
Benjamin Taura	KPSK-TUM-
Gideon Getange	KPSK-TUM-026
Collins Mambea	KPSK-TUM-08

1. Linearity between the Actual and Predicted Value in G3

The scatter plot indicates a weak correlation between actual and predicted values of G3, implying that the predictive model may need improvement.

2. Normality

In this plot, the points generally follow the red line, indicating that the data is approximately normally distributed.

3. Homoscedasticity

The variance of the residuals changes with different levels of predicted G3 values.

4. Multicollinearity

The variables in the model have a low multicollinearity with other predictors. This is evident in Medu, Fedu, famrel, freetime, goout, Dalc, health, absences with a value of (1.) .

Walc (2.033048): A VIF slightly above 2 suggests moderate correlation.

1. Write down the final, evaluate and validate model.

The final model based on the regression as per the results, can be expressed as an equation. The equation will include the significant predictors along with their coefficients. In our case, the significant predictors are "Medu" (Mother's Education) and "goout" (Going Out with Friends). Here's how we will write the final model equation:

$$G3 = 8.4165 + 0.8619 \times \text{Medu} - 0.7661 \times \text{goout}$$

This equation indicates that the predicted final grade (G3) is equal to the intercept (8.4165) plus the coefficient of "Medu" (0.8619) multiplied by the mother's education level, minus the coefficient of "goout" (0.7661) multiplied by the frequency of going out with friends.

Note, this model should be interpreted cautiously as it explains only a small portion of the variability in the final grades and might not capture all relevant factors influencing academic performance. In addition, statistically not significant predictors are excluded (Fedu, famrel, freetime, Dalc, Walc, health, absences) from the model to simplify interpretation and improve model parsimony.