a. Interpret the final model coefficients.

```
Residuals:
Min 1Q Median 3Q Max
-308.21 -5.25 -1.64 1.60 1416.82
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.920e+00 3.133e-01 22.086 < 2e-16 ***
X2.0.1 1.646e-02 3.091e-03 5.326 1.01e-07 ***
X2.0.2 2.573e-01 7.365e-03 34.938 < 2e-16 ***
X2.0.4 2.039e-02 4.479e-03 4.552 5.32e-06 ***
X0.0.15 5.513e-01 1.098e-01 5.020 5.18e-07 ***
X10.0.1 -1.641e-01 7.363e-03 -22.285 < 2e-16 ***
X0.0.20 2.409e-04 3.802e-05 6.335 2.39e-10 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 33.07 on 52389 degrees of freedom
Multiple R-squared: 0.231, Adjusted R-squared: 0.2309
```

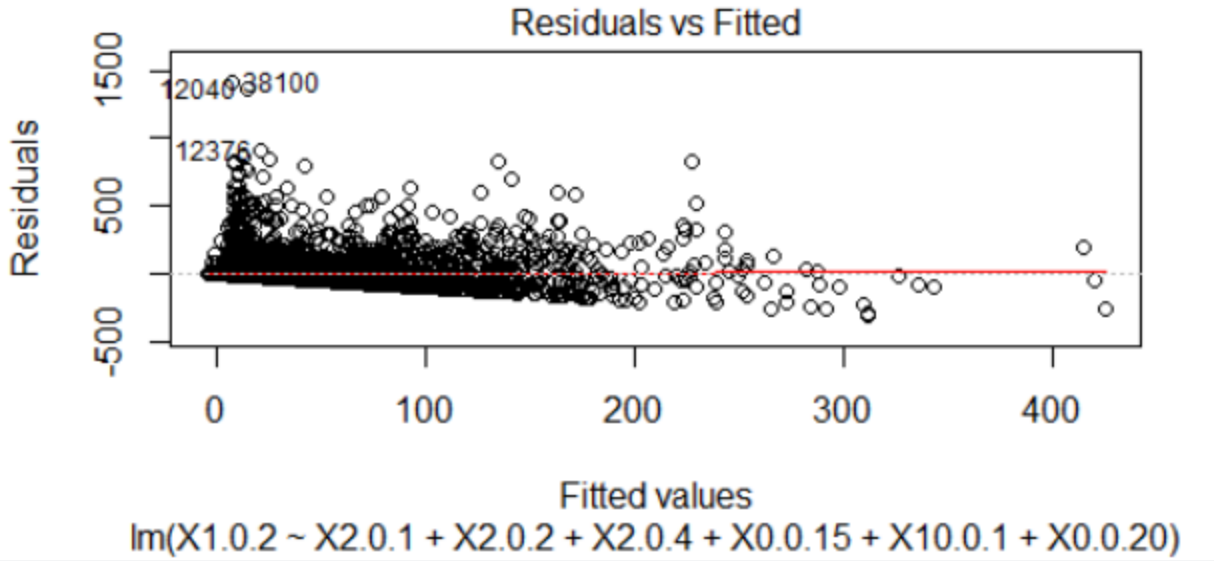
F-statistic: 2623 on 6 and 52389 DF, p-value: < 2.2e-16

- R squared value is around 23 % Which signifies that 23 % of variance in number of comments in 24 hrs is explained by this model.
- Against each coefficient we have the hypothesis testing and p value for test Null hypothesis = model coefficient is equal to 0. As p values are lower than alpha level we can conclude that model coefficient are not 0 (alternate hypothesis).
- Residual standard error signifies how far apart are the observed and predicted values.
- The Intercept is the value of Y when all Xs are 0.
- The other coefficients are values of multpiple xs in Y = ax1 + bx2 = cx3 …….

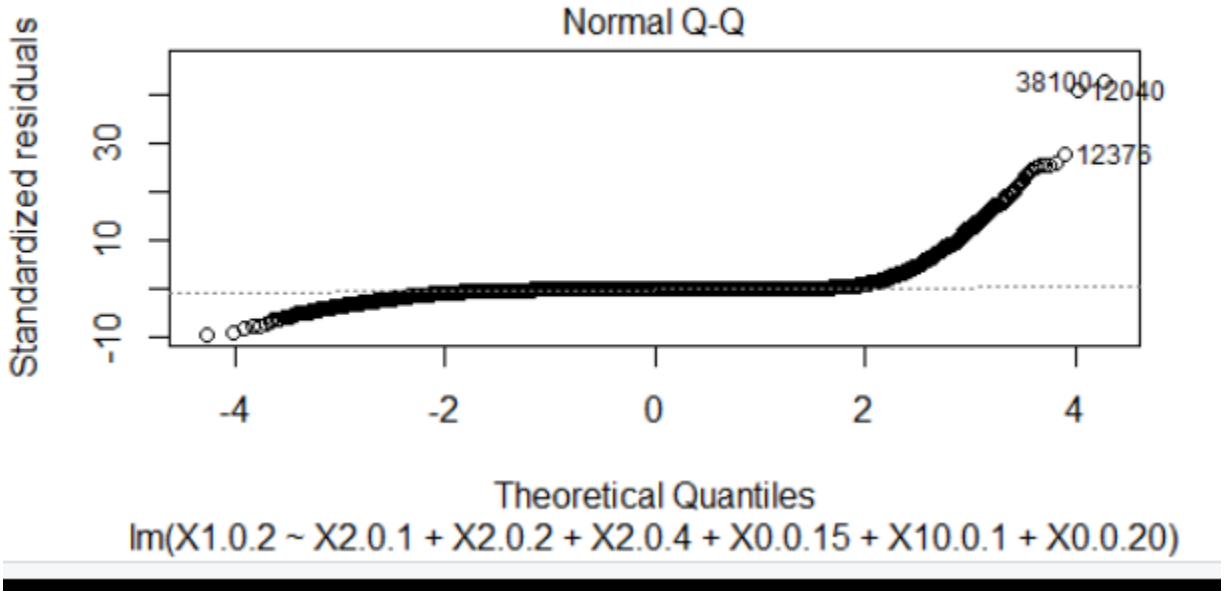
- b. Plot the model result and compare it with assumptions of the model.

<u>ASSUMPTIONS</u>

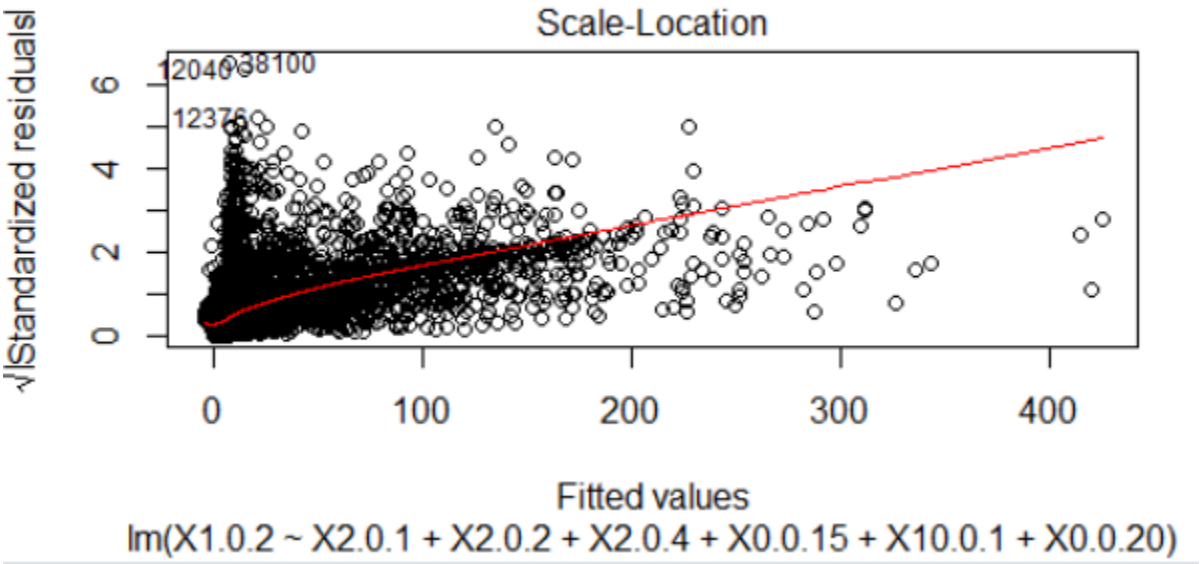
- Y values are linearly related to the X values
- Variation of observations around the regression line is constant.
- Distribution is Normal.



In the above plot we see that the line is flat which indicates a good linear relationship.



For the above plot a diagonal line would have been a perfect linear fit. However the other is not diagonal which indicates a need for improvement on the linearity of the model.



`