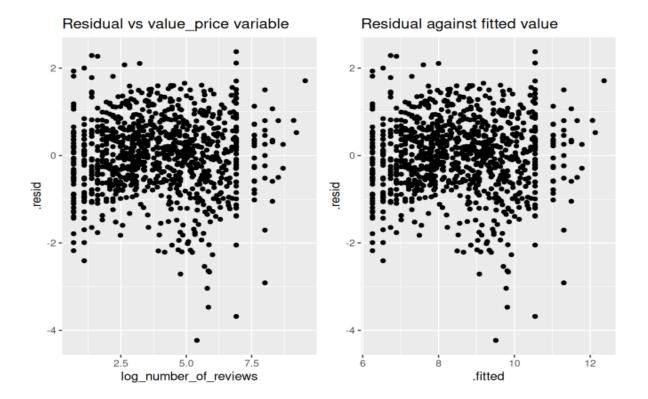**Appendix L**

L.I.N.E. assumptions analysis for Log number of reviews variable

Linearity Assumption for log number of review **variable:**

Residual vs. predictor and Residual vs. fitted values plots reveal a U-shaped pattern, suggesting a potential violation of the linearity assumption. This observation suggests that the relationship between Log love and Log number of review may not be linear. Therefore, caution should be exercised when interpreting the results of linear regression models built with this type of variable.

```{r}
# fit linear model
love_lm3 <- lm(log_love ~ log_number_of_reviews, data = sephoraData)

# Augment LR for love
love_lm3_aug <- augment(love_lm3)

# plot the residuals vs the predictors
p1 <- ggplot(data = love_lm3_aug, aes(y = .resid,
      x = log_number_of_reviews)) +
      geom_point() +
      ggtitle("Residual vs value_price variable")

# plot the residuals vs the fitted values
p2 <- ggplot(data = love_lm3_aug, aes(y = .resid, x = .fitted)) +
      geom_point() +
      ggtitle("Residual against fitted value")

plot_grid(p1, p2, nrow = 1)

```

Plots of residual vs. Log number of review and plot of residual vs. Fitted value

Independence assumption for log number of reviews variable:

The residual vs. order plot displays a scattered distribution of residuals, suggesting no discernible pattern or trend. This random sequence sustains the independence assumption between log love response and log number of reviews variable predictor, indicating no violation of this assumption in the regression model.

```{r}
# Plot residual vs. time or order
love_lm3_aug <- love_lm3_aug %>%
      mutate(order = seq(1:dim(love_lm3_aug)[1]))

ggplot(love_lm3_aug, aes(x = order, y = .resid)) +
      geom_point() +
      theme_minimal() +
      ggtitle("redisual vs. order")
```

Residual vs. order for the number of reviews variable