Question 1.2: No Perfect Collinearity

DATASCI 203 Homework: Classical Linear Model Practice

Wesley Chang

Evalute the No perfect collinearity assumption:

We want to evaluate the condition of no perfect collinearity for the model:

$$\ln(\text{views}) = \beta_0 + \beta_1 \text{rate} + \beta_3 \text{length}$$

This is fairly easy to evaluate computationally as R will automatically drop the variables that are perfectly correlated with another variable in the model.

```
summary(lm(log(views) ~ rate + length, data = videos))
```

```
##
## Call:
## lm(formula = log(views) ~ rate + length, data = videos)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.6037 -1.2665 -0.0152  1.2506  6.6876
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.411e+00  4.410e-02 122.697  < 2e-16 ***
## rate        4.725e-01  1.038e-02  45.498  < 2e-16 ***
## length      4.680e-04  7.779e-05   6.016 1.85e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.798 on 9606 degrees of freedom
##   (9 observations deleted due to missingness)
## Multiple R-squared:  0.1896, Adjusted R-squared:  0.1894
## F-statistic:  1124 on 2 and 9606 DF,  p-value: < 2.2e-16
```

We can see that the original full model is retained with no variables dropped, therefore we can safely assume there is no perfect collinearity. Additionally, we can see that the standard errors associated with each variable are fairly low, so we can further assume there is likely also no near perfect collinearity.