

modeling

```
library(GGally)
library(ggplot2)
library(lmtest)
library(lubridate)
library(moments)
library(sandwich)
library(stargazer)
library(tidyverse)
library(data.table)

source('./functions/get_robust_se.R')
source('./functions/get_clean_dataset.R')

data_clean <- get_clean_dataset()
glimpse(data_clean)

## Rows: 5,408
## Columns: 22
## $ installs      <dbl> 500, 500, 500, 1000, 1000, 1000, 1000, 1000, ~
## $ size          <dbl> 22.0, 10.0, 2.5, 5.4, 2.4, 13.0, 6.5, 53.0, 15.0, ~
## $ reviews        <dbl> 156, 120, 124, 726, 112, 112, 123, 787, 141, 171, ~
## $ rating         <dbl> 4.6, 4.5, 4.8, 4.8, 4.5, 4.4, 4.8, 4.8, 5.0, 4.8, ~
## $ price          <dbl> 3.49, 0.00, 0.00, 3.99, 7.99, 0.00, 0.00, 1.99, 0.~
## $ is_free         <lgl> FALSE, TRUE, TRUE, FALSE, FALSE, TRUE, TRUE, FALSE~
## $ last_updated   <dbl> 0.402739726, 0.010958904, 1.958904110, 0.632876712~
## $ android_ver    <dbl> 4.4, 5.0, 4.2, 4.1, 4.0, 4.4, 3.0, 4.0, 4.1, 2.2, ~
## $ current_version <dbl> 1.1, 1.5, 1.2, 2.3, 3.0, 4.5, 3.1, 1.4, 1.8, 1.7, ~
## $ category        <chr> "MEDICAL", "DATING", "FAMILY", "BUSINESS", "PRODUC~
## $ is_family_category <lgl> FALSE, FALSE, TRUE, FALSE, FALSE, FALSE, FALSE, FA~
## $ is_game_category  <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, F~
## $ is_tools_category <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, F~
## $ genre            <chr> "Medical", "Dating", "Entertainment", "Business", ~
## $ content_rating   <chr> "Everyone", "Mature 17+", "Everyone", "Everyone", ~
## $ is_content_everyone <lgl> TRUE, FALSE, TRUE, TRUE, TRUE, TRUE, TRUE, T~
## $ install_group    <int> 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~
## $ log_installs     <dbl> 2.69897, 2.69897, 2.69897, 3.00000, 3.00000, 3.000~
## $ log_size          <dbl> 1.3424227, 1.0000000, 0.3979400, 0.7323938, 0.3802~
## $ log_current_version <dbl> 0.3222193, 0.3979400, 0.3424227, 0.5185139, 0.6020~
## $ log_last_updated   <dbl> 0.146977097, 0.004733502, 0.471130891, 0.212953395~
## $ log_reviews        <dbl> 2.193125, 2.079181, 2.093422, 2.860937, 2.049218, ~

model_small <- lm(log_installs ~ 1 + rating, data = data_clean)
model_medium <- lm(log_installs ~ 1 + rating + log_size + log_current_version +
                     log_last_updated + is_free + is_family_category +
                     is_game_category + is_tools_category + is_content_everyone,
                     data = data_clean)
model_large <- lm(log_installs ~ 1 + rating + log_size + log_current_version +
```

```

    log_last_updated + is_free + is_content_everyone +
    rating * is_family_category + rating * is_game_category +
    rating * is_tools_category,
  data = data_clean)

stargazer(
  model_small,
  model_medium,
  model_large,
  type = "text",
  se = list(get_robust_se(model_small), get_robust_se(model_medium),
            get_robust_se(model_large))
)

## -----
## ====== Dependent variable: log_installs ======
## -----
## (1)          (2)          (3)
## -----
## rating      0.448***   0.338***   0.288***  

##             (0.032)     (0.031)     (0.039)  

##  

## log_size    0.432***   0.429***  

##             (0.033)     (0.033)  

##  

## log_current_version 0.430***   0.428***  

##             (0.057)     (0.057)  

##  

## log_last_updated -0.909*** -0.879***  

##             (0.071)     (0.071)  

##  

## is_free     1.418***   1.423***  

##             (0.050)     (0.051)  

##  

## is_family_category 0.032      0.384  

##             (0.037)     (0.333)  

##  

## is_game_category  0.624***  -2.374***  

##             (0.046)     (0.450)  

##  

## is_tools_category 0.272***  -0.125  

##             (0.054)     (0.423)  

##  

## rating:is_family_category -0.085  

##             (0.080)  

##  

## rating:is_game_category  0.702***  

##             (0.106)  

##  

## rating:is_tools_category 0.095  

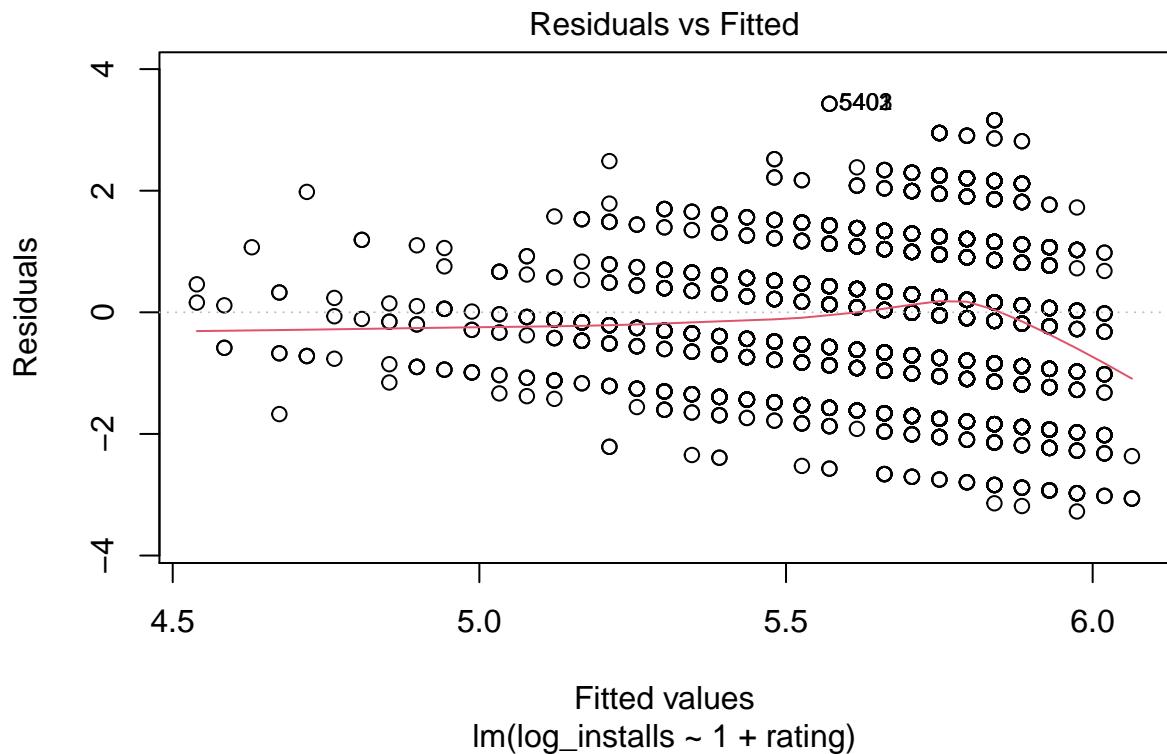
##             (0.104)

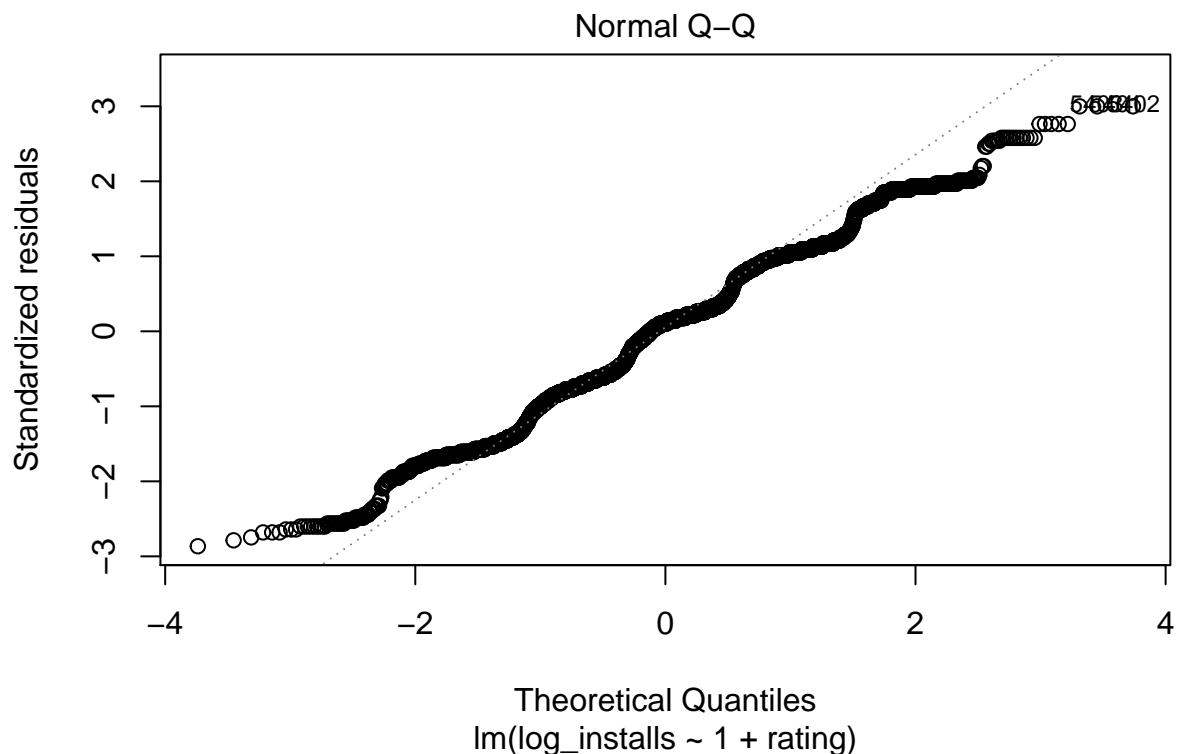
```

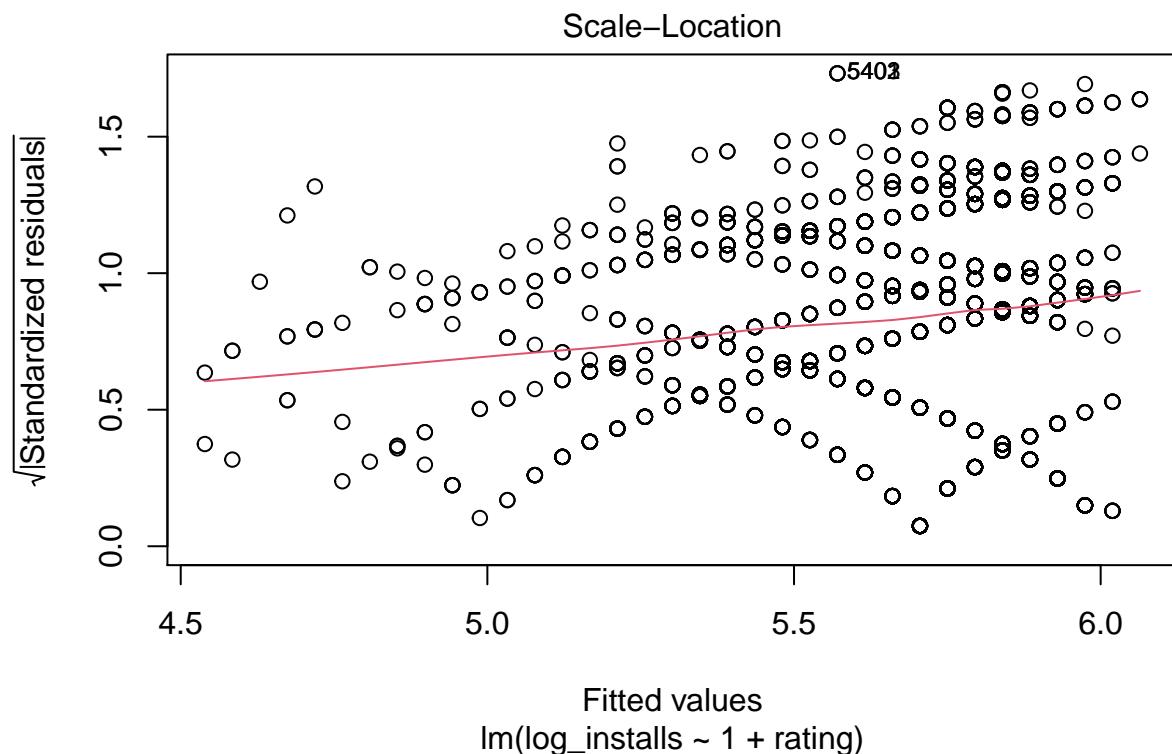
```

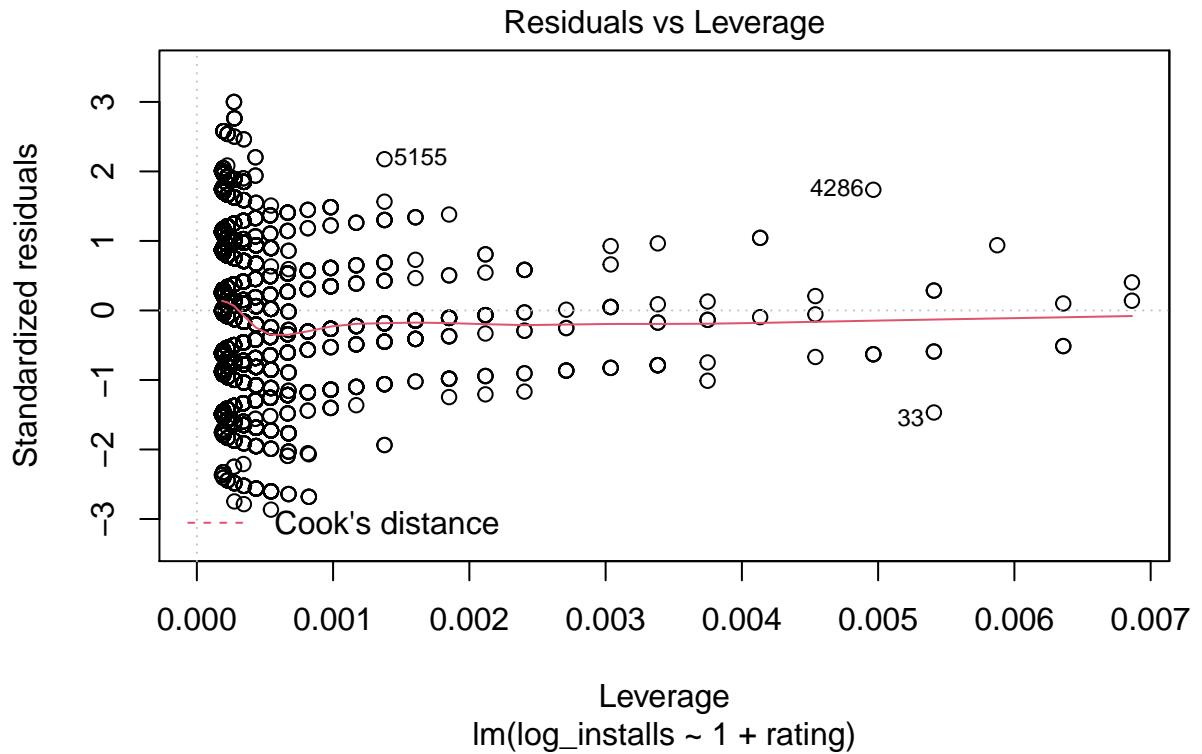
## is_content_everyone           -0.067*
##                                         (0.035)          -0.066*
##                                         (0.035)
## Constant                     3.822***      2.261***      2.469*** 
##                                         (0.133)          (0.154)          (0.181)
## -----
## Observations                  5,408          5,408          5,408
## R2                           0.028          0.265          0.271
## Adjusted R2                   0.028          0.264          0.269
## Residual Std. Error          1.143 (df = 5406) 0.995 (df = 5398) 0.991 (df = 5395)
## F Statistic                  155.421*** (df = 1; 5406) 216.073*** (df = 9; 5398) 166.929*** (df = 12; 5395)
## Note:                         *p<0.1; **p<0.05; ***p<0.01
plot(model_small)

```









```
plot(model_medium)
```

