Final models and Stargazer table with Confirmation data

Roshni Tajnekar, Erik Sambrailo, Jenna Sparks, Bailey Kuehl

2023-04-13

```
##     revenue              budget              votes            rating
##  Min.   :2.019e+05   Min.   :   100000   Min.   :   2000   Length:379
##  1st Qu.:2.500e+07   1st Qu.: 12000000   1st Qu.:  35000   Class :character
##  Median :7.813e+07   Median : 30000000   Median :  74000   Mode  :character
##  Mean   :1.920e+08   Mean   : 51870923   Mean   : 131741
##  3rd Qu.:2.275e+08   3rd Qu.: 63500000   3rd Qu.: 163500
##  Max.   :2.798e+09   Max.   :356000000   Max.   :1000000
##     runtime        country
##  Min.   : 81.0   Length:379
##  1st Qu.:100.0   Class :character
##  Median :110.0   Mode  :character
##  Mean   :112.1
##  3rd Qu.:122.0
##  Max.   :209.0
```

Final models for regression

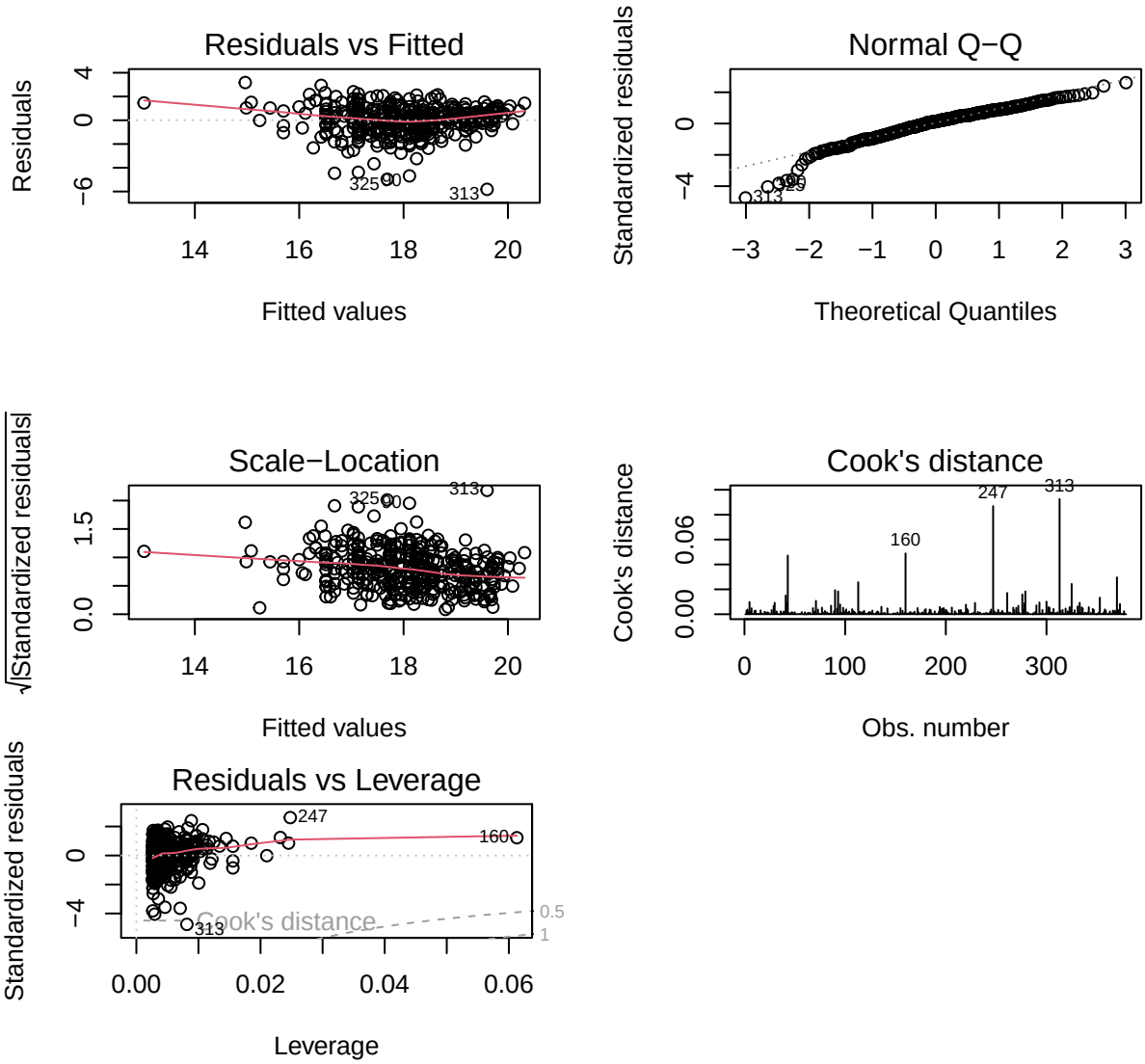
Model1 - With log(gross), log(budget)

$$ln(\text{revenue}) = \beta_0 + \beta_1 \cdot ln(\text{budget})$$

```
## (Intercept) log(budget)
##   2.7588381   0.8918258
```

```
##
## Call:
## lm(formula = log(revenue) ~ log(budget), data = conf_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.8166 -0.7073  0.1540  0.8210  3.1733
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.75884    0.90751    3.04  0.00253 **
## log(budget)  0.89183    0.05277   16.90  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.23 on 377 degrees of freedom
## Multiple R-squared:  0.4311, Adjusted R-squared:  0.4296
## F-statistic: 285.6 on 1 and 377 DF,  p-value: < 2.2e-16
```

```
## [1] 2.37357
```



Model 2 - With log(revenue), log(budget), log(votes), rating

$$ln(\text{revenue}) = \beta_0 + \beta_1 \cdot ln(\text{budget}) + \beta_2 \cdot ln(\text{votes}) + \beta_3 \cdot \text{rating}$$

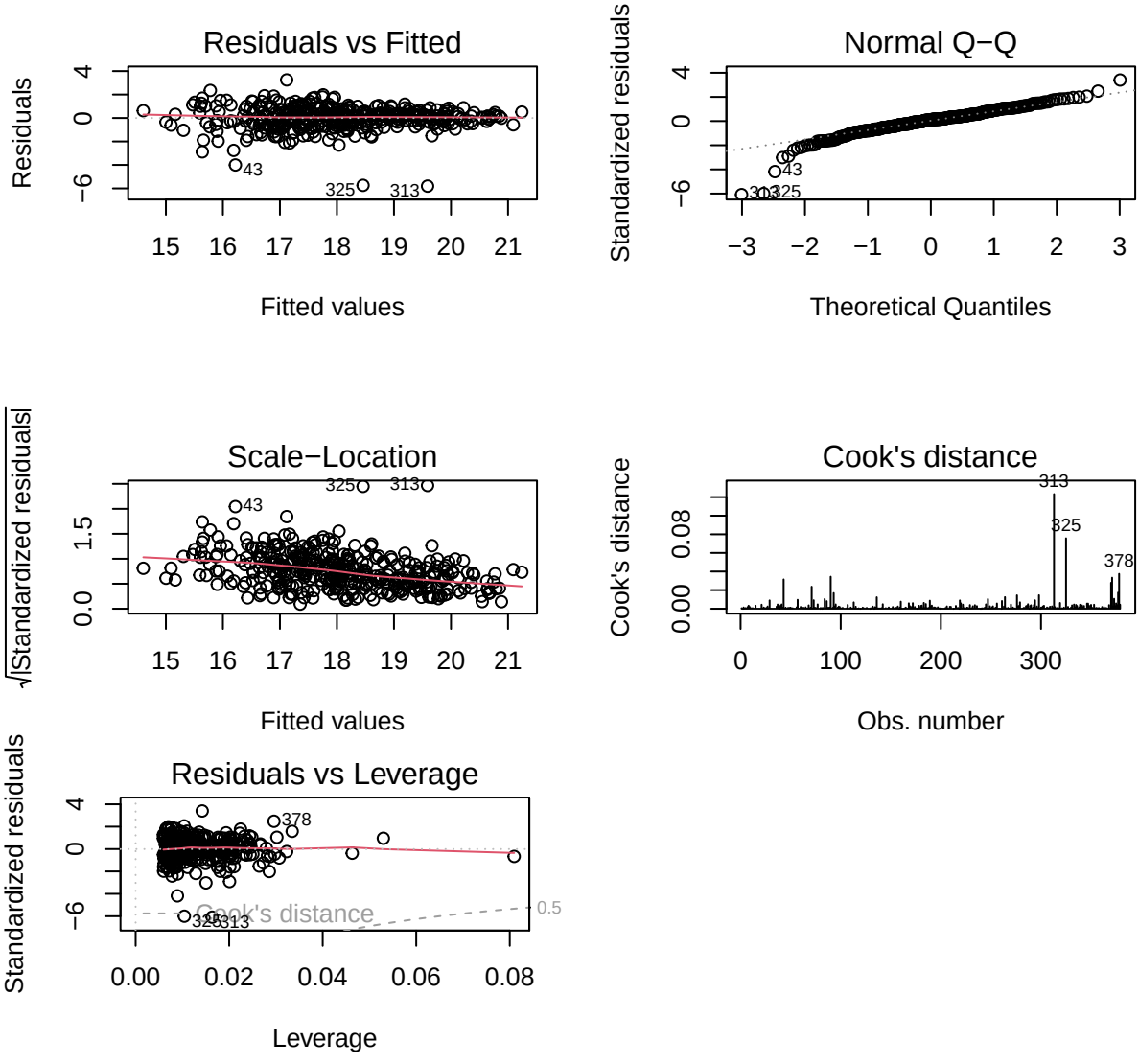
```
## (Intercept) log(budget)  log(votes) ratingPG-13     ratingR
##   2.8092082   0.4326846   0.7781838  -0.7581934  -1.3117381
```

```
##
## Call:
## lm(formula = log(revenue) ~ log(budget) + log(votes) + rating,
##     data = conf_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.8052 -0.4688  0.0945  0.5333  3.2478
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.80921    0.80109   3.507 0.000509 ***
## log(budget)  0.43268    0.05149   8.403 9.19e-16 ***
```

```
## log(votes)   0.77818    0.05171  15.049  < 2e-16 ***
## ratingPG-13 -0.75819    0.15710  -4.826 2.03e-06 ***
## ratingR     -1.31174    0.16483  -7.958 2.11e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9619 on 374 degrees of freedom
## Multiple R-squared:  0.6546, Adjusted R-squared:  0.6509
## F-statistic: 177.2 on 4 and 374 DF,  p-value: < 2.2e-16
```

```
##                 GVIF Df GVIF^(1/(2*Df))
## log(budget) 1.556043  1        1.247415
## log(votes)  1.413510  1        1.188911
## rating      1.296482  2        1.067067
```

```
## [1] 7.093795
```



Model 3 - With log(revenue), log(budget), log(votes), rating, runtime, country

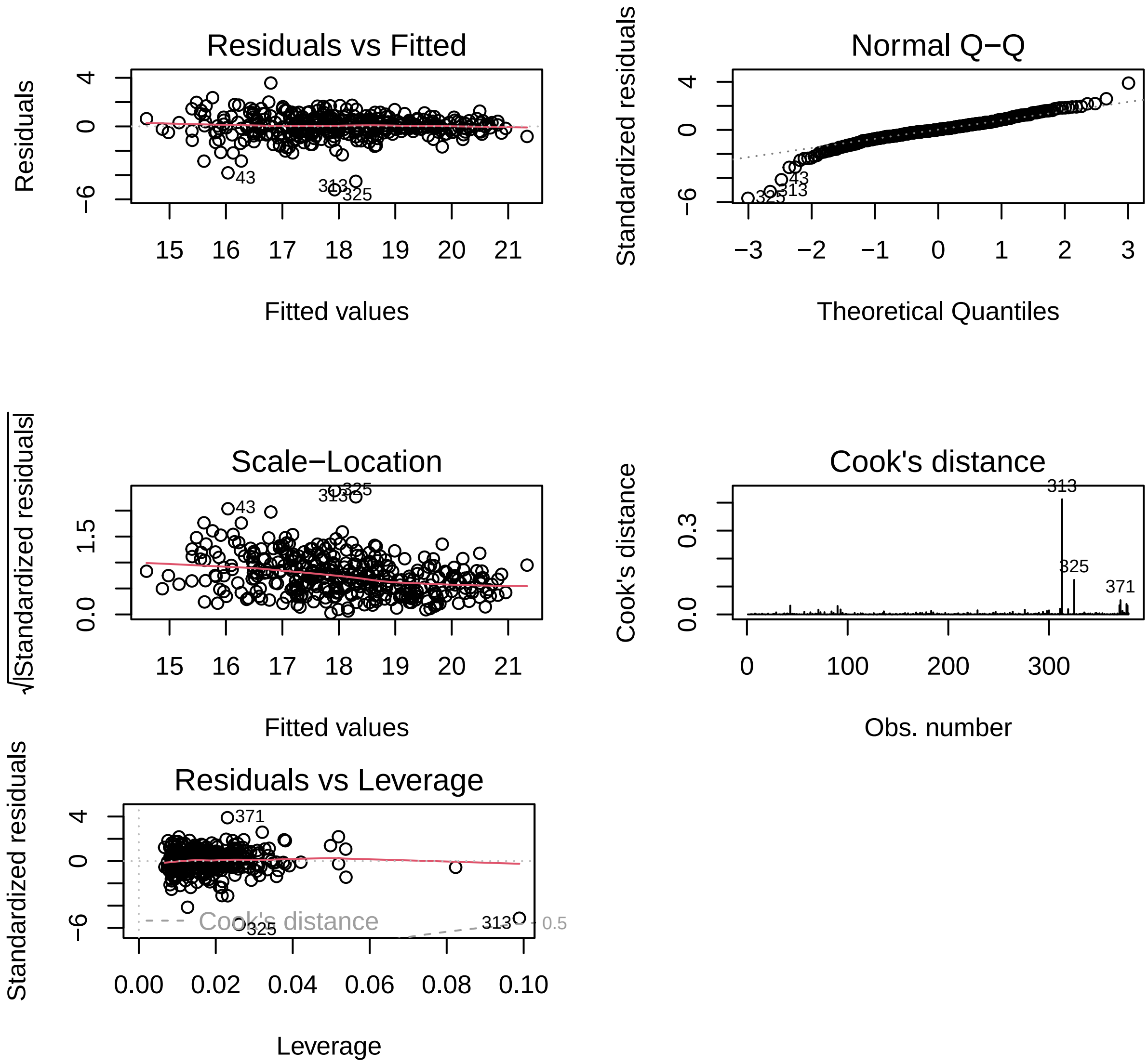
$$ln(\text{revenue}) = \beta_0 + \beta_1 \cdot ln(\text{budget}) + \beta_2 \cdot ln(\text{votes}) + \beta_3 \cdot \text{age certification} + \beta_4 \cdot \text{runtime} + \beta_5 \cdot \text{country}$$

```
##          (Intercept)          log(budget)           log(votes)
##           2.34082703           0.52621728           0.81242056
##          ratingPG-13              ratingR              runtime
##          -0.58857165          -1.11534499          -0.01645173
## countryUnited States
##           0.21984819
```

```
##
## Call:
## lm(formula = log(revenue) ~ log(budget) + log(votes) + rating +
##     runtime + country, data = conf_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.2069 -0.4485  0.0298  0.5005  3.5714
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           2.34083    0.77780   3.010 0.002795 **
## log(budget)           0.52622    0.05289   9.950  < 2e-16 ***
## log(votes)            0.81242    0.05108  15.906  < 2e-16 ***
## ratingPG-13          -0.58857    0.15646  -3.762 0.000196 ***
## ratingR              -1.11535    0.16498  -6.761 5.33e-11 ***
## runtime              -0.01645    0.00329  -5.000 8.83e-07 ***
## countryUnited States  0.21985    0.11393   1.930 0.054401 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9281 on 372 degrees of freedom
## Multiple R-squared:  0.6801, Adjusted R-squared:  0.675
## F-statistic: 131.8 on 6 and 372 DF,  p-value: < 2.2e-16
```

```
## [1] 4.887043
```

```
##                 GVIF Df GVIF^(1/(2*Df))
## log(budget) 1.762934  1        1.327755
## log(votes)  1.481193  1        1.217043
## rating      1.401778  2        1.088103
## runtime     1.383156  1        1.176077
## country     1.026124  1        1.012978
```



Stargazer table with above models

```
##
## OLS Models of Revenue based on Budget, votes and ratings for movie revenue empirical study
## ======================================================================
##                                      Dependent variable:
##                          ---------------------------------------------------
##                                          log(revenue)
##                              (1)              (2)              (3)
## ----------------------------------------------------------------------
## log(budget)              0.892***         0.433***         0.526***
##                          (0.056)          (0.055)          (0.055)
##
## log(votes)                                0.778***         0.812***
##                                           (0.058)          (0.061)
##
## ratingPG-13                              -0.758***        -0.589***
##                                           (0.122)          (0.144)
##
```

```
## ratingR                                -1.312***       -1.115***
##                                         (0.157)         (0.156)
##
## runtime                                                 -0.016**
##                                                         (0.006)
##
## countryUnited States                                     0.220
##                                                         (0.130)
##
## Constant                 2.759**         2.809**         2.341**
##                          (0.981)         (0.874)         (0.805)
##
## ------------------------------------------------------------------------
## Observations               379             379             379
## R2                        0.431           0.655           0.680
## Adjusted R2               0.430           0.651           0.675
## Residual Std. Error  1.230 (df = 377) 0.962 (df = 374) 0.928 (df = 372)
## ========================================================================
## Note:                                  *p<0.05; **p<0.01; ***p<0.001
```