Final models and Stargazer table with Exploratory data

Roshni Tajnekar, Erik Sambrailo, Jenna Sparks, Bailey Kuehl

2023-04-13

```
##     revenue              budget              votes           rating
##  Min.   :4.570e+04   Min.   :  1200000   Min.   :   195   Length:163
##  1st Qu.:3.651e+07   1st Qu.: 15500000   1st Qu.: 38500   Class :character
##  Median :9.328e+07   Median : 35000000   Median : 76000   Mode  :character
##  Mean   :2.318e+08   Mean   : 62071779   Mean   :138830
##  3rd Qu.:2.545e+08   3rd Qu.: 90000000   3rd Qu.:195000
##  Max.   :2.048e+09   Max.   :321000000   Max.   :897000
##     runtime        country
##  Min.   : 76.0   Length:163
##  1st Qu.: 98.0   Class :character
##  Median :110.0   Mode  :character
##  Mean   :111.3
##  3rd Qu.:121.0
##  Max.   :164.0
```

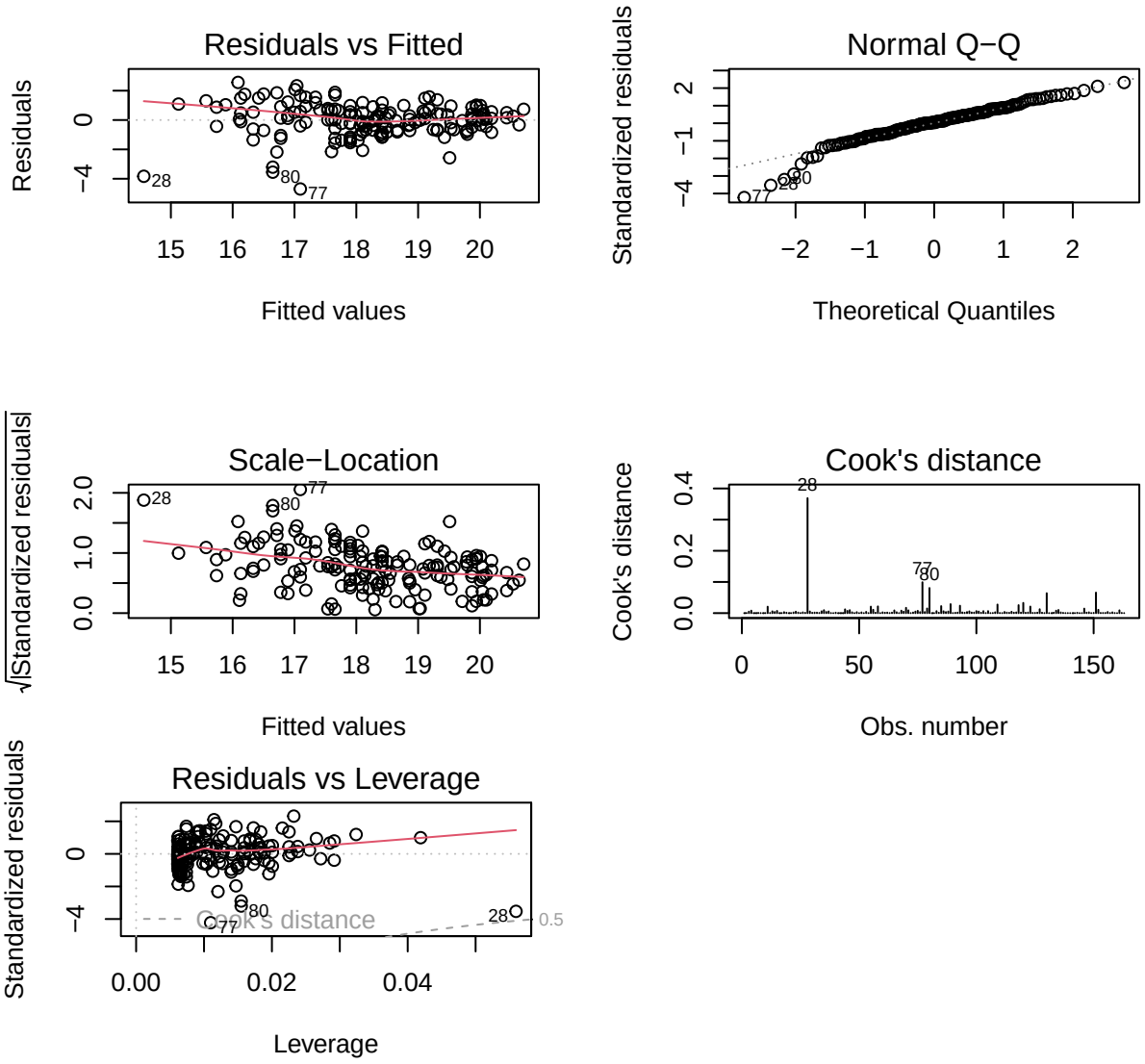
Final models for regression

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

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

```
## (Intercept) log(budget)
##  -0.8315335   1.0997121
```

```
##
## Call:
## lm(formula = log(revenue) ~ log(budget), data = explo_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6971 -0.6475  0.0528  0.6687  2.5611
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.83153    1.29593  -0.642    0.522
## log(budget)  1.09971    0.07455  14.752   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.118 on 161 degrees of freedom
## Multiple R-squared:  0.5748, Adjusted R-squared:  0.5721
## F-statistic: 217.6 on 1 and 161 DF,  p-value: < 2.2e-16
```



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

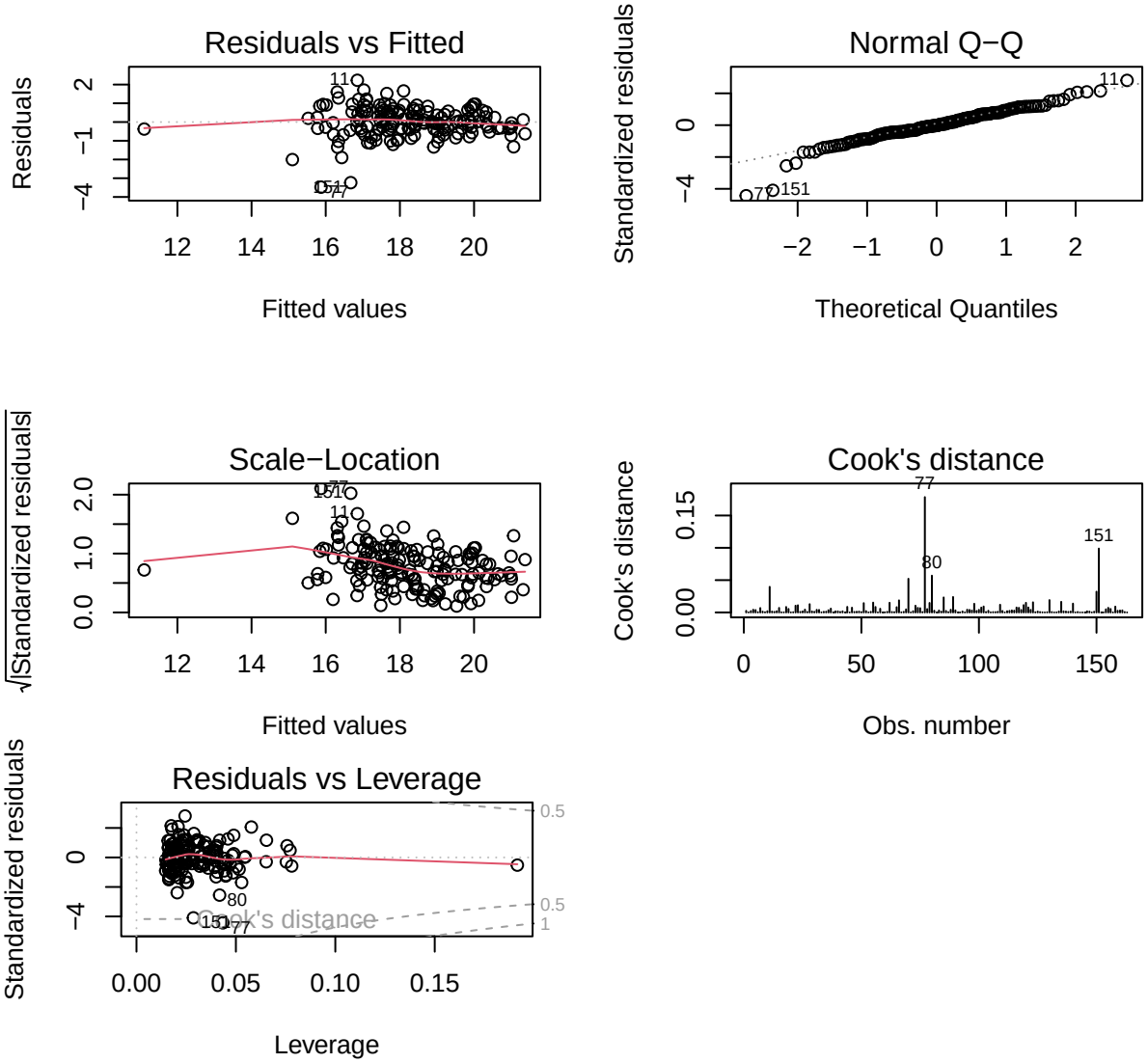
$$ln(\text{revenue}) = \beta_0 + \beta_1 \cdot ln(\text{budget}) + \beta_2 \cdot ln(\text{votes}) + \beta_3 \cdot \text{rating}$$

```
## (Intercept) log(budget)  log(votes) ratingPG-13     ratingR
##   1.7962450   0.4513205   0.8513826  -0.9825751  -1.4981750
```

```
##
## Call:
## lm(formula = log(revenue) ~ log(budget) + log(votes) + rating,
##     data = explo_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4764 -0.3702 -0.0165  0.5420  2.2272
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.79625    0.99554   1.804   0.0731 .
## log(budget)  0.45132    0.07593   5.944 1.72e-08 ***
```

```
## log(votes)   0.85138    0.07484  11.376  < 2e-16 ***
## ratingPG-13 -0.98258    0.17745  -5.537 1.26e-07 ***
## ratingR     -1.49818    0.18740  -7.994 2.58e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8019 on 158 degrees of freedom
## Multiple R-squared:  0.7852, Adjusted R-squared:  0.7797
## F-statistic: 144.4 on 4 and 158 DF,  p-value: < 2.2e-16
```

```
##                 GVIF Df GVIF^(1/(2*Df))
## log(budget) 2.015295  1        1.419611
## log(votes)  1.907831  1        1.381242
## rating      1.176959  2        1.041575
```



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

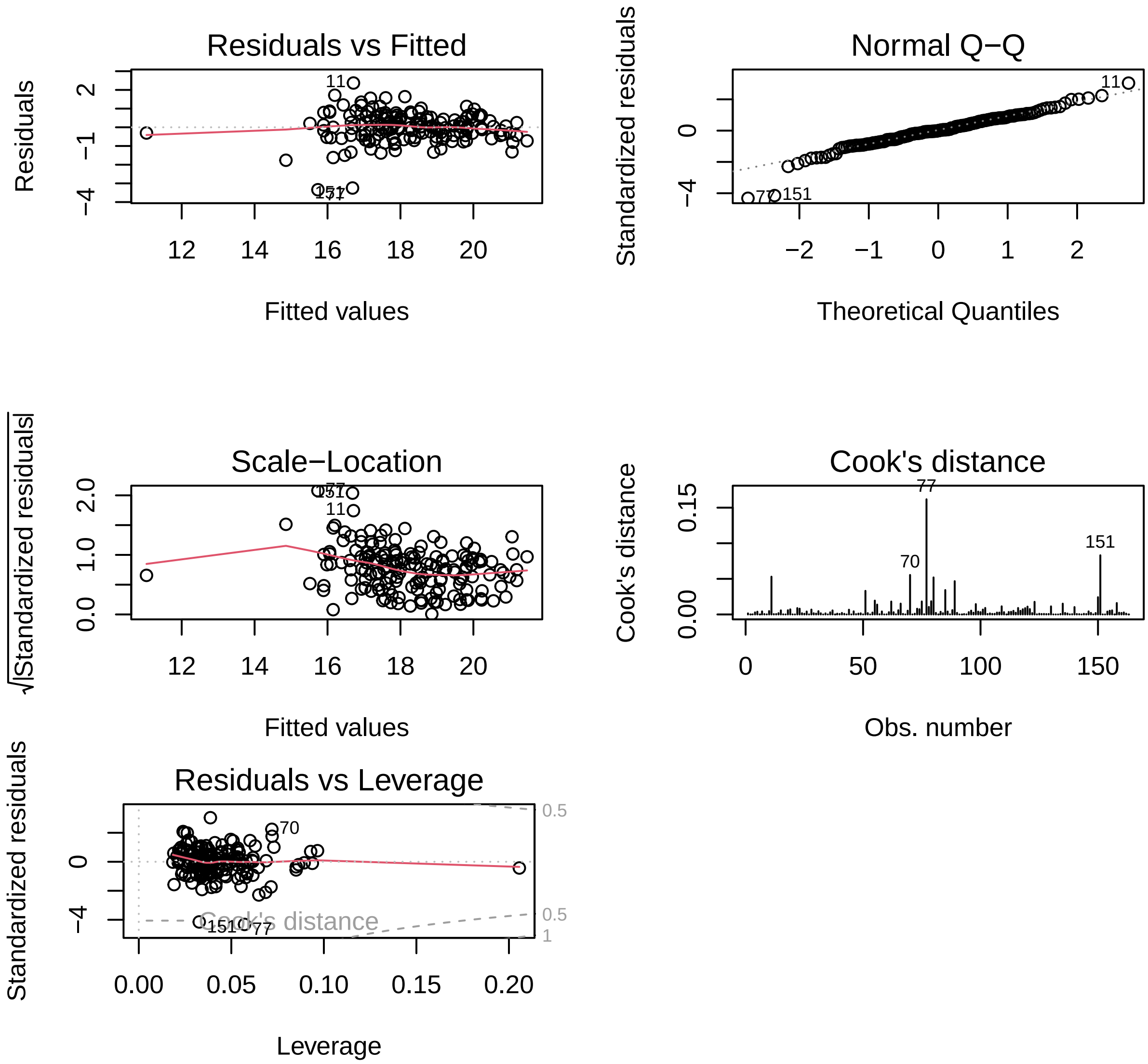
$$ln(\text{revenue}) = \beta_0 + \beta_1 \cdot ln(\text{budget}) + \beta_2 \cdot ln(\text{votes}) + \beta_3 \cdot \text{age certification} + \beta_4 \cdot \text{runtime} + \beta_5 \cdot \text{country}$$

```
##          (Intercept)          log(budget)          log(votes)
```

```
##           1.62544547           0.48318071           0.89613041
##          ratingPG-13              ratingR              runtime
##          -0.88555988           -1.37827707           -0.00931482
## countryUnited States
##           0.09075486
```

```
##
## Call:
## lm(formula = log(revenue) ~ log(budget) + log(votes) + rating +
##     runtime + country, data = explo_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3394 -0.4369  0.0000  0.5017  2.3692
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           1.625445   0.991598   1.639    0.103
## log(budget)           0.483181   0.077670   6.221 4.35e-09 ***
## log(votes)            0.896130   0.077974  11.493  < 2e-16 ***
## ratingPG-13          -0.885560   0.183038  -4.838 3.12e-06 ***
## ratingR              -1.378277   0.195376  -7.054 5.31e-11 ***
## runtime              -0.009315   0.004819  -1.933    0.055 .
## countryUnited States  0.090755   0.135876   0.668    0.505
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7958 on 156 degrees of freedom
## Multiple R-squared:  0.7911, Adjusted R-squared:  0.7831
## F-statistic: 98.49 on 6 and 156 DF,  p-value: < 2.2e-16
```

```
##                 GVIF Df GVIF^(1/(2*Df))
## log(budget) 2.141460  1        1.463373
## log(votes)  2.103110  1        1.450210
## rating      1.308987  2        1.069631
## runtime     1.627700  1        1.275813
## country     1.021721  1        1.010802
```



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)                1.100***         0.451***         0.483***
##                            (0.096)          (0.068)          (0.071)
##
## log(votes)                                  0.851***         0.896***
##                                             (0.080)          (0.089)
##
## ratingPG-13                                -0.983***        -0.886***
##                                             (0.173)          (0.182)
##
```

```
## ratingR                                -1.498***        -1.378***
##                                         (0.174)          (0.192)
##
## runtime                                                   -0.009
##                                                          (0.005)
##
## countryUnited States                                      0.091
##                                                          (0.146)
##
## Constant                    -0.832          1.796           1.625
##                            (1.716)         (1.078)         (1.124)
##
## ------------------------------------------------------------------------
## Observations                 163             163             163
## R2                          0.575           0.785           0.791
## Adjusted R2                 0.572           0.780           0.783
## Residual Std. Error  1.118 (df = 161) 0.802 (df = 158) 0.796 (df = 156)
## ========================================================================
## Note:                                  *p<0.05; **p<0.01; ***p<0.001
```