Project#1

Nathan Robinson, . . .

02/19/2020

Regression

This is a group project, and students should work in a group of size 3. Include all the R code, hypothesis testing, one or two lines of explanation for any output. The report should be organized, printed,and stapled. The due date of this project is **Wednesday** 02/19/2020.

```
Movies = read.csv("C3 2008Movies.csv")
```

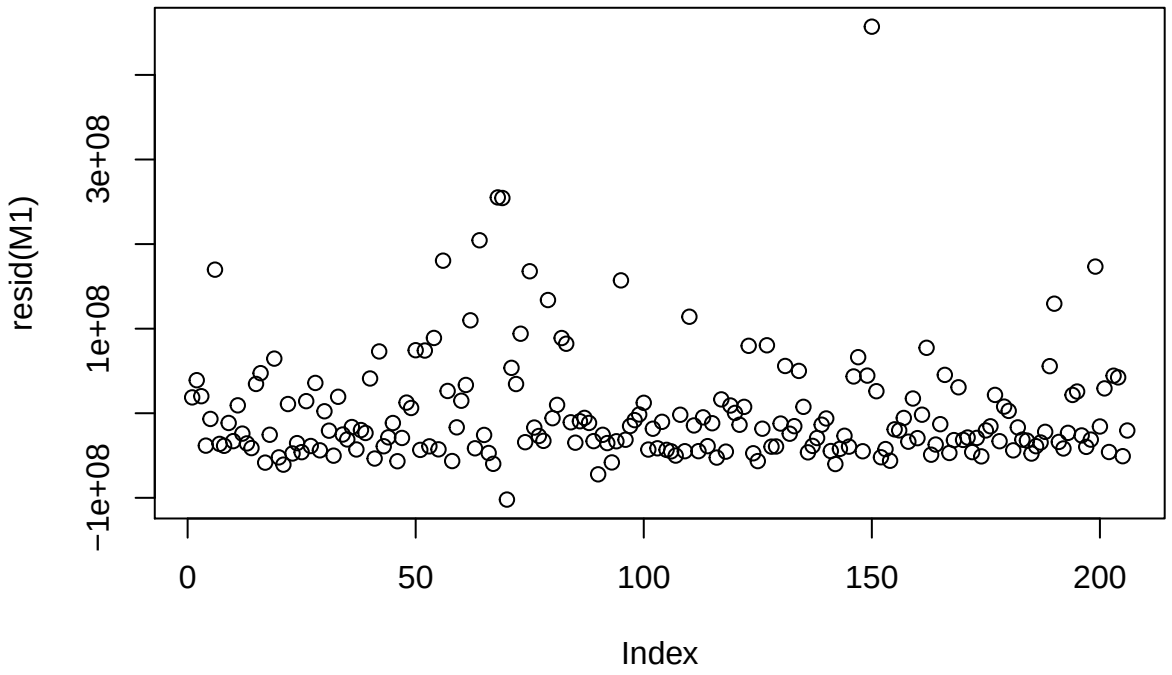
The 2008Movies file contains data on movies released in 2008.

1. Calculate a regression model to predict box office from run time. Interpret the R^2 value and test statistic for the slope in the context of this problem.

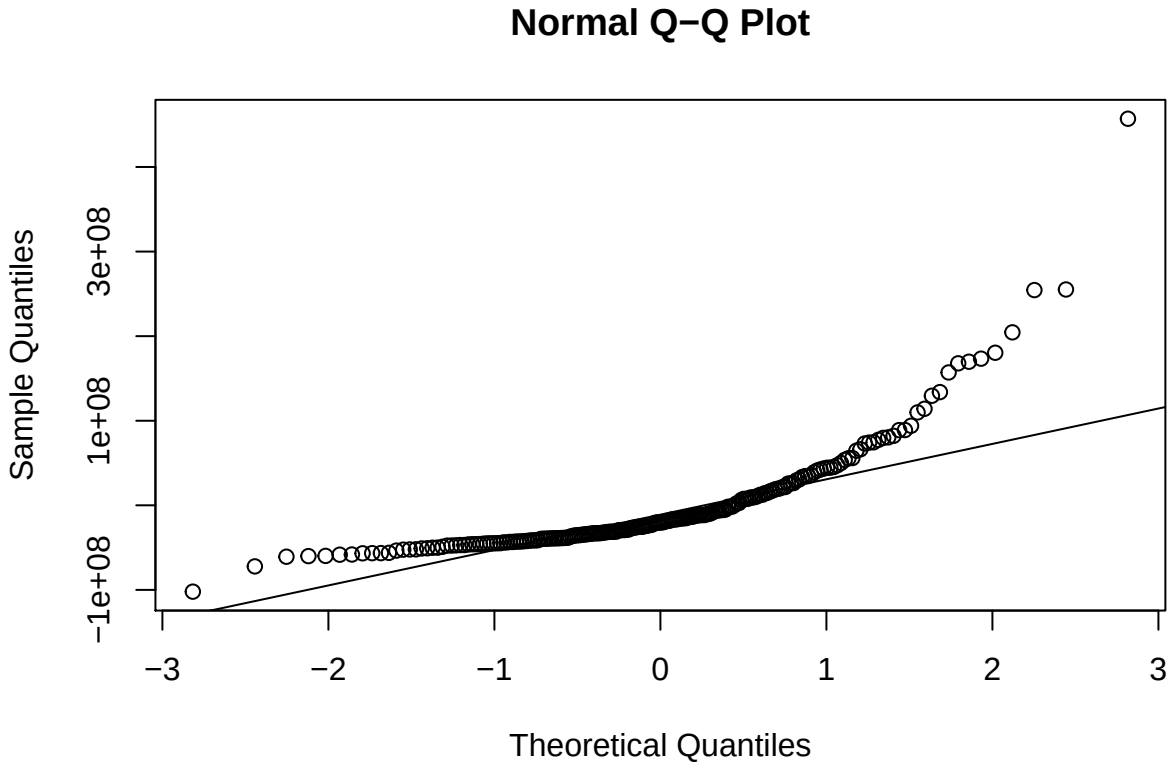
```
M1 = lm(BoxOfficeGross~RunTime, data=Movies)
summary(M1)
```

```
##
## Call:
## lm(formula = BoxOfficeGross ~ RunTime, data = Movies)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -102059739  -39266026  -20290622   17164421  457025023
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3506911   24316122   0.144    0.885
## RunTime       478843     226856   2.111    0.036 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 65930000 on 204 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.02137,    Adjusted R-squared:  0.01658
## F-statistic: 4.455 on 1 and 204 DF,  p-value: 0.03601
```

```
plot(resid(M1))
```




```
qqnorm(resid(M1))
qqline(resid(M1))
```



2. Create indicator variables for the genre and MPAA rating. Use the best subsets regression to determine a appropriate regression model.

```r
Genre = as.numeric(Movies$Genre)
MPAA = as.numeric(Movies$MPAA)

library("leaps")
Model_subset = regsubsets(BoxOfficeGross~Genre + MPAA, data=Movies)
summary(Model_subset)
```

```
## Subset selection object
## Call: regsubsets.formula(BoxOfficeGross ~ Genre + MPAA, data = Movies)
## 14 Variables  (and intercept)
##                           Forced in Forced out
## GenreAdventure                FALSE      FALSE
## GenreComedy                   FALSE      FALSE
## GenreConcert/Performance      FALSE      FALSE
## GenreDocumentary              FALSE      FALSE
## GenreDrama                    FALSE      FALSE
## GenreHorror                   FALSE      FALSE
## GenreMusical                  FALSE      FALSE
## GenreRomantic Comedy          FALSE      FALSE
## GenreThriller/Suspense        FALSE      FALSE
## GenreWestern                  FALSE      FALSE
## MPAANot Rated                 FALSE      FALSE
## MPAAPG                        FALSE      FALSE
## MPAAPG-13                     FALSE      FALSE
## MPAAR                         FALSE      FALSE
## 1 subsets of each size up to 8
## Selection Algorithm: exhaustive
##          GenreAdventure GenreComedy GenreConcert/Performance GenreDocumentary
## 1  ( 1 ) "*"            " "         " "                       " "
## 2  ( 1 ) "*"            " "         " "                       " "
## 3  ( 1 ) "*"            " "         " "                       " "
## 4  ( 1 ) "*"            " "         " "                       "*"
## 5  ( 1 ) "*"            " "         " "                       "*"
## 6  ( 1 ) " "            "*"         " "                       "*"
## 7  ( 1 ) " "            "*"         "*"                       "*"
## 8  ( 1 ) " "            "*"         "*"                       "*"
##          GenreDrama GenreHorror GenreMusical GenreRomantic Comedy
## 1  ( 1 ) " "        " "         " "          " "
## 2  ( 1 ) " "        " "         " "          " "
## 3  ( 1 ) "*"        " "         " "          " "
## 4  ( 1 ) "*"        " "         " "          " "
## 5  ( 1 ) "*"        " "         "*"          " "
## 6  ( 1 ) "*"        " "         " "          "*"
## 7  ( 1 ) "*"        " "         " "          "*"
## 8  ( 1 ) "*"        "*"         " "          "*"
##          GenreThriller/Suspense GenreWestern MPAANot Rated MPAAPG MPAAPG-13
## 1  ( 1 ) " "                    " "          " "           " "     " "
## 2  ( 1 ) " "                    " "          " "           " "     " "
## 3  ( 1 ) " "                    " "          " "           " "     " "
## 4  ( 1 ) " "                    " "          " "           " "     " "
## 5  ( 1 ) " "                    " "          " "           " "     " "
## 6  ( 1 ) "*"                    " "          " "           " "     " "
## 7  ( 1 ) "*"                    " "          " "           " "     " "
```

```
## 8  ( 1 ) "*"                    " "           " "           " "     " "
##           MPAAR
## 1  ( 1 ) " "
## 2  ( 1 ) "*"
## 3  ( 1 ) "*"
## 4  ( 1 ) "*"
## 5  ( 1 ) "*"
## 6  ( 1 ) "*"
## 7  ( 1 ) "*"
## 8  ( 1 ) "*"
```

```
# We want row 8
# GenreComedy + GenreConcert/Performance + GenreDocumentary + GenreDrama + GenreHorror + GenreRomantic
```

a. Validate the model assumptions.

b. Look at residual plots and check for heteroskedasticity (unequal variance), multicollinearity, correl

c. submit your suggested least squares regression formula along with a limited number of appropriate gra

d. Test the overall model adequcy.

3. Conduct an extra sum of squares test to determine if one or more interaction terms (or quadratic terms) should be included in the model. You can choose any other terms to test.
4. Test whether average run time is the same for different Genre. Clearly show your hypothesis test.
5. Check equality of variance of run time for Genre type.