STAT 3110 Applied Regression (Fall 2020)

Project (20 points)
Due Dec. 20, 2020 11:59pm

Directions:

- You will need to submit a write-up that summarizes your findings and also includes your SAS codes and necessary SAS outputs.
- Submit your write-up electronically via Canvas, in either word or pdf format.
- Name your write-up as "LastName-FirstName-StudentID".
- The write-up must be clean, well-organized and easily readable (3 points).

Predicting Movie Grosses. The file named MOVIES can be found on Canvas. It contains data on movies released in the United States during the calendar year 1998. The variables in this file are

- TDOMGROSS: the total domestic gross revenue (dependent variable)
- BACTOR: the number of "best actors" in the movie; to qualify as a best actor, the person must be listed in *Entertainment Weekly*'s lists of the 25 Best Actors and 25 Best Actresses of the 1990s
- TDACTOR: the number of actors or actresses appearing in the movie who were among the top 20 actors and top 20 actresses in average box office gross per movie in their careers ("top dollar actors"), according to The Movie Times website at the beginning of the 1998 movie season
- GENRE: classification of movie types coded as 1 = action, 2 = drama, 3 children's, 4 = comedy, 5 = documentary, 6 = thriller, 7 = horror, 8 = science fiction

- MPAA: MPAA rating coded as 1 = G, 2 = PG, 3 = PG13, 4 = R, 5 = NC17, 6 = unrated
- COUNTRY: where the movie was made coded as 1 = USA, 2 = English-speaking country other than USA, 3 = non-English speaking country
- CHRISTMAS: coded as 1 if the movie was to be released during the Christmas season, 0 otherwise
- HOLIDAY: coded as 1 if the movie was released before any holiday weekend (President's Day, Memorial Day, Labor Day, Independence Day, Labor Day, Thanksgiving, or the Christmas season), 0 otherwise
- SUMMER: coded as 1 if the movie was released during the summer (Memorial Day through Labor Day), 0 otherwise
- SEQUEL: coded as 1 if the movie was a sequel, 0 otherwise

We would like to find an equation to predict the total domestic gross revenue of movies based on variables available prior to the release of the films. All of the variables in the data file are known or could be determined prior to release. Use regression to examine the relationships and determine what factors might be useful in helping to predict gross revenues. Use the log of TDOMGROSS as the dependent variable. Some of the explanatory variables are qualitative in nature and will need to be transformed into indicators. What relationships do you find that would be helpful in predicting revenues? What variables are not important? What recommendations would you make on the basis of your results?

These data are dicussed in the article "Prediction Movie Grosses: Winners and Losers, Blockbusters and Sleepers," by Jeffrey S. Simonoff and Ilana R. Sparrow, *Chance*, Vol. 13, 2000, 15-24 and were obtained from Dr. Simonoff's website.