

STA 302 H1F / 1001 HF – Fall 2011 – A. Gibbs

Assignment 2

Do “better” movies make more money?

Due: Thursday, November 10 at 12:00 noon.

Include both your name and student number on your assignment. Underline your surname.

Your assignment should be handed in at lecture on that day. Assignments will be accepted before lecture starts, at the break, or at the end of lecture. (Please don't interrupt the lecture if you come in late.) No late assignments will be accepted without a valid reason. If you cannot attend lecture on that day, make arrangements to hand in the assignment *before* lecture.

Presentation of solutions is important. In particular, it is inappropriate to hand in pages of SAS output without explanation or interpretation. The only SAS output you need to submit with assignments is relevant plots. Quote relevant numbers from your SAS output as part of your solutions. You do not need to hand in your SAS code.

The Data:

The data for this question are available on the course web site under Assignments or at www.utstat.utoronto.ca/alison/Teaching/1112/Sta302/assignments.html.

You can read about the study from which the data are extracted at

http://www.usatoday.com/life/movies/2004-02-25-cover-critics-fans_x.htm.

The data are from movies released in 2003 that had wide distribution. We are interested in whether good reviews lead to better earnings. The variables you are given are:

- Critics' score, out of 100. This is an average calculated from movie reviews published in 20 major American newspapers, magazines and wire services.
- How much money the movie made at the box office in millions of dollars.
- The movie's rating (G, PG, PG-13, or R). You can read about the ratings system at <http://www.filmratings.com>.

If you're curious what the movies are, their names are available in another file in the same order as in the data file.

1. Consider a scatterplot of the data as given. What features of the plot indicate that a transformation of the dependent variable is necessary? Your answer should reference concerns about the fit of the model or a violation of the Gauss-Markov conditions that a transformation might improve.
2. In lecture we discussed three (square root, natural log, inverse) of the many possible transformations that are possible. Based on plots, pick the best model among these three choices for a transformation of the dependent variable. Justify your choice.
3. For the model you chose in question 2, interpret the estimated slope in a practical way. Your answer should be in terms of the original variables, not transformations of them.
4. Do you have any concerns about the model you chose in question 2? In addition to plots, consider appropriate statistics that we have discussed in class.
5. Although we won't fully investigate the role of rating in determining a movie's box office earnings, you can check whether or not rating seems to have an effect by looking at a scatterplot

of the transformed data (using the transformation you chose in question 2), with different symbols for different ratings. What do you learn from the plot and what do you think you should do? Your answer should not be independent of your answer to question 4.

Here is some suggested SAS code:

```
symbol1 v=circle c=black;  
symbol2 v=star c=black;  
symbol3 v=triangle c=black;  
symbol4 v=square c=black;
```

```
proc gplot;  
  plot y*x=rating;
```

6. The methodology of the original study was described in *USA Today*. They say that they examined a model using all of the movies you are given, as well a model fit to the data with the top-grossing movies excluded. Top-grossing movies were defined as those that earned more than \$200 million at the box office. Fit the model you chose in question 2 to the data with the top-grossing movies excluded. Do you think the top-grossing movies should be excluded from the analysis? Why or why not?

Marking scheme: Each of the questions is worth 3 marks (for a total of 18). 3 marks will be awarded for complete, correct answers, or answers with only very minor problems. Good answers that are unclear or have some mistakes or are missing some aspects of the solution will be awarded 2 marks. Poor answers that have some value will be awarded one mark. Note that sometimes an answer awarded 3 marks will not be perfect. You should always look at the solutions.