Statistics 104 - Term Project

Due: December 3, 2021, correction Nov 24

The file *hollywoodmovies.xls* contains information on movies from 2011. Analyze the data set with respect to the following variables:

1. Estimate the center of the distribution and 95% confidence interval for the following variables:
   (a) budget, domestic gross product, opening weekend calculating means and t-intervals.
   (b) compare analysis of the mean using t-intervals to medians and associated intervals for medians.
2. Compare budget, domestic gross and opening weekend of comedies to dramas. Use two-sample t-tests and compare your results to non-parametric equivalents. Which method should be used. Justify your answer with plots and other statistical assessments.
3. Compare budget, domestic gross and opening weekend for the following three categories:
   (a) horror/thriller
   (b) **animation/fantasy/romance**
   (c) action/adventure. Use analysis of variance based on the assumption of normality and non-parametric equivalents. State the assumptions for analysis and assess which if any are violated.
4. For action movies, compare domestic gross vs opening weekend. Are the intakes correlated? Calculate the following correlations: Pearson and Spearman Rank correlation. Which method should be used based on the assumptions required. Find the regression of domestic gross on budget. Again, use parametric and non-parametric methods. Which method is preferred and why?
5. Compare domestic to foreign gross for action movies. Where are earnings higher, inside or outside the US? Use an appropriate parametric test and compare to the non-parametric equivalent.

In all cases, assess the distribution your data is sampled from using plots and other diagnostic tools. Discuss which approach you prefer and why. Also, discuss various non-parametric tests and which one you feel is the most appropriate for each of the samples. Summarize your results in the form of written report not to exceed 5 pages, in the style similar to a scientific article. You will be graded on both the statistical methods used and the write-up. The appendix is NOT part of the 5 page write-up. You can work alone or with one other person from class.

Your article should consist of the following parts

1. An abstract
2. An introduction/overview
3. A methods/results section which presents your analysis in detail
4. A discussion section which elaborates on your methods/results section
5. An appendix that contains all computer code used and evidence that you actually used the computer code.