HW2: Baseball big bang?
(assigned on Mon April 10, due by 4pm on Wed April 12)

You may work with in a group of as many as three students on this assignment, handing in one report with all names, provided that you all contribute to the work. You must submit a word-processed report, with computer output integrated into your report as appropriate.

A reader wrote in to the “Ask Marilyn” column in Parade magazine to say that his grandfather told him that in 3/4 of all baseball games, the winning team scores more runs in one inning than the losing team scores in the entire game. (This phenomenon is known as a “big bang.”) Marilyn responded that this proportion seemed to be too high to be believable. I wanted to test the grandfather’s hypothesis against Marilyn’s alternative for Major League Baseball (MLB) games in the 2017 season, so I selected a sample of games and recorded whether or not a “big bang” occurred in each game.

a) Identify the observational units and variable in this study. Also classify the variable as categorical (also binary?) or quantitative.

b) Identify the population of interest in this study. Also describe the relevant parameter (in words).

c) Convert the grandfather’s assertion into a null hypothesis, using symbols.

d) Report Marilyn’s response as an alternative hypothesis, in symbols and in words.

To investigate these hypotheses, I examined the 36 MLB games played on April 2-5, 2017, the first few days of the season.

e) Does this constitute a simple random sample of MLB games played in the 2017 season? Explain why or why not.

I found that 19 of these 36 games contained a big bang.

f) Calculate the sample proportion of games that had a big bang, and denote it with the appropriate symbol.

g) Use the One Proportion Inference applet to investigate whether the sample data provide strong evidence against the grandfather’s claim in support of Marilyn’s alternative. Specify the input values that you enter for the probability of success, sample size, and number of samples. Also submit a screen capture of the resulting distribution of sample proportions.

h) Report the (approximate) p-value from your simulation analysis.

i) Would you reject the null hypothesis at the $\alpha = .05$ significance level?
j) Based on the observed sample data and the simulation results, would you conclude that the sample data provide strong evidence to support Marilyn’s contention that the proportion cited by the grandfather is too high to be the actual value? Explain your reasoning, as if writing to the grandfather, who has never taken a statistics course.

k) Now consider the null hypothesis that 40% of all 2017 MLB games contain a big bang. Conduct a simulation analysis of this hypothesis, using a two-sided alternative. Submit a screen capture of the distribution of sample proportions, and report the (approximate) p-value. Also summarize your conclusion.