Homework 2

The Challenger Launch

Due date: Wednesday, November 4th, 2020.

This homework will show you how logistic regression analysis could have allowed the management of NASA to potentially decide differently when deciding whether to postpone the launch of the Space Shuttle Challenger in January 28, 1986, and thus, prevent the disaster that occurred on that day.

The background:
"Up until January 28, 1986, the Challenger had been launched 24 times with no catastrophes. A catastrophe can occur, in particular, if the collective failure of all the so-called O-rings occurs. However, 7 of the 24 Challenger launches had experienced at least one O-ring failure.
During one of such launches, three of five rings failed. One launch had two O-rings failed, and five launches had a single failure. So collectively, there were ten O-ring failures.

On the day of the January 28th launch, which had already been postponed a few times due to bad weather and other technical difficulties, the temperature at the launch site, Kennedy Space Center, Florida, was unusually cold with morning temperature close to 30 degrees Fahrenheit minus one degree Celsius. Some engineers expressed concerns about the launch in such cold weather given that there was no data to certify this successful launch at those temperatures. It was eventually decided to go ahead with the launch.

Tragically, Challenger experienced a catastrophe were all of the five O-Rings failed leading to the destruction of the shuttle and killing everybody on board. The launch and the subsequent disaster were broadcasted live on television to a large audience across the United States, making headline news."[1]

The data:
A shuttle has 5 O-rings, and there were 24 launches, which gives us a total of 120 observations. By looking at the data, one launch experienced 3 O-ring failures, one experienced 2 O-ring failures, and all the others experienced 1 O-ring failure.

Logistic regression:

1. Logistic regression analysis is a sibling of the linear regression analysis that we have studied so far. Read about logistic regression, how to perform it using R, and how to interpret the coefficients. Recommended sources: Chapter 8 of the textbook and Link .
2. Now we will turn to a real-world use of logistic regression. Make a scatter plot for the temperatures at launch and the number of O-rings that failed using data only for when

[1] Citation is hidden in order to prevent web-searches that could foster academic dishonesty.

the failures of O-rings did occur. Can you tell if there is a significant relationship between the temperatures and O-ring failures?

3. The scatterplot that you just made ignores valuable information. Now, make a scatterplot with the complete data that includes the launches where no failures occurred. Can you now tell if there is a significant relationship between these two variables?
4. We will now analyze this problem with regression analysis where Y will be a binary variable showing failure versus no-failure and X will be an independent variable showing temperature at time of the launch. Using logistic regression establish a relationship between independent and dependent variables and show the regression summary. Is temperature significant in determining the no failure versus failure of O-rings? What does the sign for the coefficient of Beta 1 tells you?
5. At the 95% significance level, what is the confidence interval for Beta 1? Can you tell if the coefficient for Beta 1 is significant just by looking at the confidence interval? How?
6. By how much do the odds of failure change when the temperature increases by one-degree Fahrenheit?
7. What was the probability of failure of all of the 5 O-rings on the morning of January 28, 1986 given that the temperature at launch was 36 degrees Fahrenheit? Explicitly, calculate in R the following:
   $$P(y = 1 \mid X = 36) = \frac{e^{\beta_0 + \beta X}}{1 + e^{\beta_0 + \beta X}}$$
8. Create a table (or a matrix) that shows the probability of failures for individual O-rings at various temperatures according to your model. Explicitly, create a table with 3 columns, one for temperature (X), another one for the probability of failure of one O-ring given the different values of temperature in the data or Pr(Y=1|X), and another one for the probability that all 5 O-rings fail or Pr(Catastrophe|X).
9. Would you have recommended the launch of the Challenger on January 28, 1986? Why? What are the limitations of your conclusions?