W271 Group Lab 1

Investigating the 1986 Space Shuttle Challenger Accident

Denny Lehman, Mingxi Liu, Aruna Bisht, Deepika Maddali

Contents

Abstract

This report will, indeed, be abstract. No, instead, describe your goals your approach, and what you learn.

1 Introduction

1.1 Research question

The purpose of this research is to reproduce and examine the results of Dalal et al.(1989). In that paper, the authors quoted the Rogers Commission report on the space shuttle Challenger accident, stating that the accident was caused by a combustion gas leak through a joint in one of the booster rockets, which was sealed by a device called an O-ring. The paper collected data from the 23 pre-accident launches and built both binomial and binary logistic regression models to examine the influence of launching temperature and leak test pressure on the probability of O-ring incidents. It concluded that there was strong statistical evidence of a temperature effect on incidents of O-ring while the pressure effect was not significant.

This leads to our research question: **"Does temperature and pressure have a significant impact on O-ring failure of space shuttles? If so, what is the implication of the these factors on the risk of the Challenger launch?"**

2 Data

2.1 Description

The data were collected from the rocket motors recovered from the launches before Challenger. There were 24 launches, but for one flight the motors were lost at sea, so motor data was available for 23 flights.

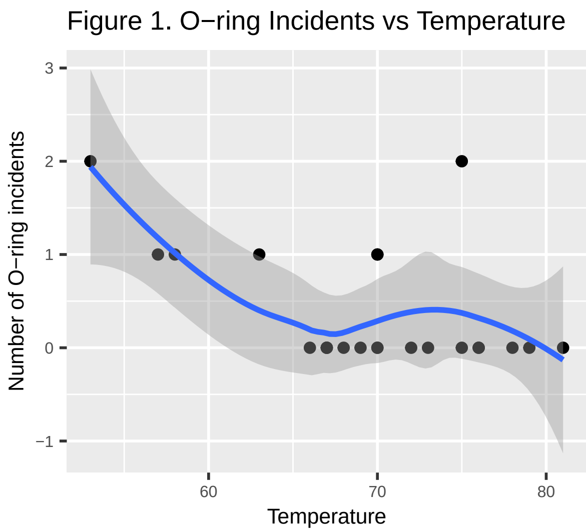
Our purpose is to find out whether or how the incidents of O-ring thermal distress, are related to launching temperature and leak-check pressure. The *O.ring* in the data indicates the number of primary O-ring incidents. Each shuttle has 6 primary O-rings, so the *number* in the data is always 6. *Temp* and *Pressure* are the temperature and pressure mentioned before, and *Flight* is just the flight ID.

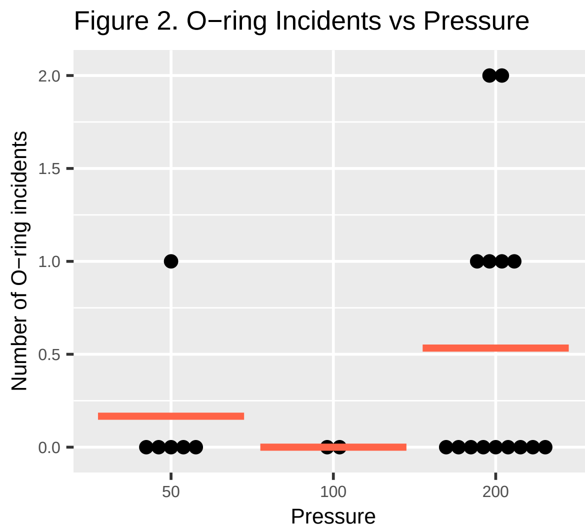
Dalal et al.(1989) built a binominal logistic regression model upon the data. This requires two assumptions of independence: 1) each of the 23 launches was independent; 2) each O-ring incident within the same launch was also independent. The first assumption, although debateble because later launches benefitted from the experience of earlier lauches, is generally acceptable. But the second assumptions requires more caution and knowledgeable. Without the second assumption, i.e. the failure of one O-ring may indicate the failures of other O-rings, we don't have 23*6 independent binomial response for each O-ring. Instead we only have 23 binary observations indicating whether there was any O-ring incidents for each launches. That's why Dalal et al.(1989) also built a binary model as a supplement to the binominal model and argued that the conclusion is robust to the second assumption.

Table 1: Table 1. Statisctics of variables

	mean	std	min	max	VIF
O.ring	0.3913	0.6564	0	2	
Temp	69.5652	7.0571	53	81	1.0016
Pressure	152.1739	68.2213	50	200	1.0016

2.2 Key Features





The data is quite simple and clean: both *O.ring* and *Pressue* have only 3 possible values and *Temp* ranges from 53 to 81 without extreme values. Both *Temp* and *Pressure* have a VIF close to 1, suggesting no multicollinearity. From Figure 1 we can observe that as temperature decreases, the number of incidents tends to increase, with one outlier. It's more obvious when we add a smoothed trendline (the blue line). The "S"-shaped trendline suggests a logistic regression may be suitable. Figure 2 suggests that *O.ring* is not likely to be monotonically related to *Pressure*, but the average number of O-ring incidents (as denoted by the red line) seems to be higher when *Pressure* is at 200. It may be worth considering including *Pressure* as a categorical variable in the model.

3 Analysis

3.1 Reproducing Previous Analysis

We attempt to reproduce the analysis from the Dalal et al. With the Challenger data described above, we create a logistic regression model to determine the relationship between O-ring failure and explanatory variables pressure and temperature. An interesting property of the pressure readings discovered in the EDA is that there are only 3 pressure readings in the data, 50, 100, and 200 psi. For this reason, we will attempt to model pressure as either a continuous or categorical variable. Modeling pressure as a category will help determine if one or more levels of pressure is significant.

$$logit(\pi) = \beta_0 + \beta_1 * temp$$

$$logit(\pi) = \beta_0 + \beta_1 * temp + \beta_2 * pressure$$

$$logit(\pi) = \beta_0 + \beta_1 * temp + \beta_2 * pressure_{100} + \beta_3 * pressure_{200}$$

```
##
## Table 2. The estimated relationship between O-ring failure an
## =====================================================
##                              Dependent variable:
##                        -------------------------------
##                                 O.ring/Number
##                           (1)        (2)         (3)
## -----------------------------------------------------
## Temp                    -0.116*    -0.098*    -0.096*
##                         (0.047)    (0.045)    (0.045)
## Pressure                            0.008
##                                    (0.008)
## factor(Pressure)100                            -15.845
##                                              (3,282.153)
## factor(Pressure)200                             1.067
##                                               (1.101)
## Constant                 5.085      2.520      2.992
##                         (3.052)    (3.487)    (3.213)
## -----------------------------------------------------
## Observations              23         23         23
## Log Likelihood          -15.823    -15.053    -14.718
## Akaike Inf. Crit.       35.647     36.106     37.435
## =====================================================
## Note:                    *p<0.05; **p<0.01; ***p<0.001
```

The stargazer table shows the three models, one with temperature as the only explanatory variable (1), one with temperature and pressure as a continuous variable (2), and one model with temperature and pressure as a categorical variable (3).

Model 1 estimates β_1 as -0.116 with a standard error of 0.047. The wald test statistic was converted to a p-value to determine if the coefficient of each explanatory variable was significant to the model. The null hypothesis for this test is that the coefficient of the explanatory variable is 0. We use a critical value of $\alpha = 0.05$.

With a Wald p-value of 0.014 which is less than the critical value $\alpha = 0.05$, we reject the null hypothesis. Temperature has a non zero effect on the model, all else equal. The intercept term on the other hand, was not statistically significant. The estimate for the intercept was 5.08 with a standard error of 3.05. With a wald p-value of 0.096 which is greater than 0.05, we fail to reject the null hypothesis that the intercept term is non zero. The interpretation of the statistically significant temperature term is for every 1 unit increase in temperature, there is a -0.116 decrease in the log odds of O.ring failure, all else equal.

For the second model, we analyze temperature and pressure as explanatory variables to logit probability of O-ring failure. The estimate for the temperature coefficient was -0.098 with a standard error of 0.045. With a p-value of 0.0285455 which is less than 0.05, we reject the null hypothesis. Temperature has a non zero effect on this model, all else being equal, just as in the first model. The estimate for $\beta_{pressure}$ is 0.008 with a standard error of 0.008. With a p-value of 0.269, we fail to reject the null hypothesis. There is not enough evidence to suggest that the coefficient for pressure is non zero. Finally, the intercept term has an estimate of 2.52 with a standard error of 3.487. With a wald p-value of 0.47 which is greater than 0.05, we fail to reject the null hypothesis that the intercept term is non zero. The only statistically significant term in model 2 is temperature. The interpretation is that for every unit increase in temperature there is a -0.098 decrease in the log odds probability of O-ring failure, all else equal.

Another way of including Pressure in the model is to treat it as a categorical variable since it only has three possible values. However, as shown in the table above, under this treatment the Pressure variable is still insignificant.

Additionally, the likelihood ratio test (LRT) was used to determine if the models were different. First we consider model 1 with temperature vs model 2 which includes continous pressure. The null hypothesis is that there is no difference between the models, such that adding additional parameters has no effect on the MLE. The alternative hypothesis is that adding additional terms improves the model performance, such that the more complex model is better.

```
## Analysis of Deviance Table
##
## Model 1: O.ring/Number ~ Temp
## Model 2: O.ring/Number ~ Temp + Pressure
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        21     18.086
## 2        20     16.546  1   1.5407   0.2145
```

With a p-value of 0.21452 which is greater than 0.05, we fail to reject the null hypothesis. The model with only temperature is the better model.

```
## Analysis of Deviance Table
##
## Model 1: O.ring/Number ~ Temp
## Model 2: O.ring/Number ~ Temp + factor(Pressure)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        21     18.086
## 2        19     15.875  2   2.2112    0.331
```

The model using pressure as a categorical variable fared just as poorly in the LRT. Comparing the temperature only model with the model using pressure as a categorical variable, the p-value of the LRT was 0.3310193. Since the p-value is larger than 0.05, we fail to reject the null hypothesis, adding pressure as a continuous or categorical variable does not improve the model.

3.2 Confidence Intervals

We consider the simplified model $logit(\pi) = \beta_0 + \beta_1 Temp$, where π is the probability of an O-ring failure. We test to determine if the model that includes an additional quadratic temperature term improves the model:

$$logit(\pi) = \beta_0 + \beta_1 Temp + \beta_2 Temp^2$$

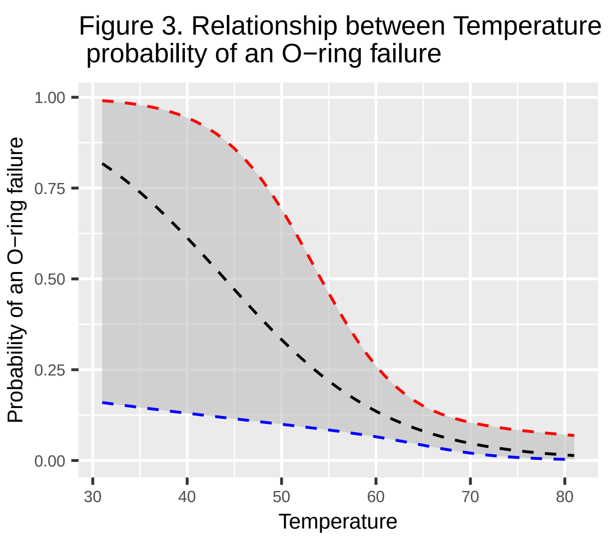
We use the anova function to perform the likelihood ratio test to determine whether the quadratic temperature term is associated with O-ring failure probability. The results will determine if $Temp^2$ should be included in the model.

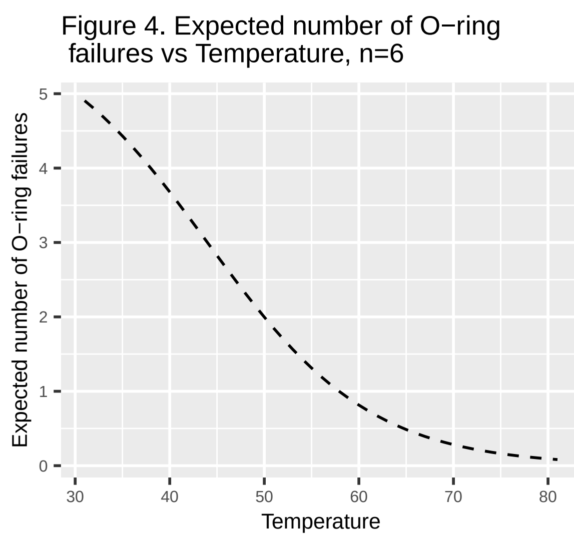
- $H_0 : \beta_2 = 0$
- $H_a : \beta_2 \neq 0$

```
## Analysis of Deviance Table
##
## Model 1: O.ring/Number ~ Temp
## Model 2: O.ring/Number ~ Temp + I(Temp^2)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        21     18.086
## 2        20     17.592  1   0.4947   0.4818
```

With a p-value of 0.482 which is greater than the critical value $\alpha = 0.05$, we fail to reject the null hypothesis. There is not enough evidence to suggest that adding the quadratic temperature term improves the logistic model, all else being equal.

The original model performs best, so we calculate the 95% confidence intervals around the estimated probability of O-ring failure, $\widehat{pi}$, for all temperature intervals between 31° F to 80° F. Below are a plot of the results





Given this model, the estimated probability of an O-ring failure at the launch temperature of 31°, the same as the Challenger launch in 1986, is 0.8178 with a 95% confidence interval of (0.1596, 0.9907) for each O-ring.

** ADD explaination on inference here!!!! **

3.3 Bootstrap Confidence Intervals

In the previous section, we employed the Wald method to construct a confidence interval, which relies on asymptotic theory. Wald's method assumes large sample sizes and specific assumptions, such as normality and independence, to provide reliable estimates. However, in our case with only 23 observations which is a small sample size, these assumptions may not hold true, and the Wald method may lead to inaccurate or unreliable confidence intervals.

To overcome this limitation, we adopted the parametric bootstrap approach. This approach is a resampling technique that is not heavily reliant on large sample sizes but instead assumes that the data follows a particular parametric distribution. Specifically, we used the logistic parametric model, which is suitable for modeling discrete response variables like in our case the occurrence of O-ring failure in our flight dataset.

The first step in the parametric bootstrap approach was to simulate new datasets. Using the sample function, we randomly selected observations from the original dataset of 23 observations and replaced them, allowing for the same observation to be selected multiple times. This resampling process enables us to create new datasets with a distribution similar to the original dataset. We then fitted a logistics regression model to each bootstrap resampled dataset, estimating the regression coefficients and other model parameters. With these fitted models, we predicted the probability of ring failure at each temperature point.

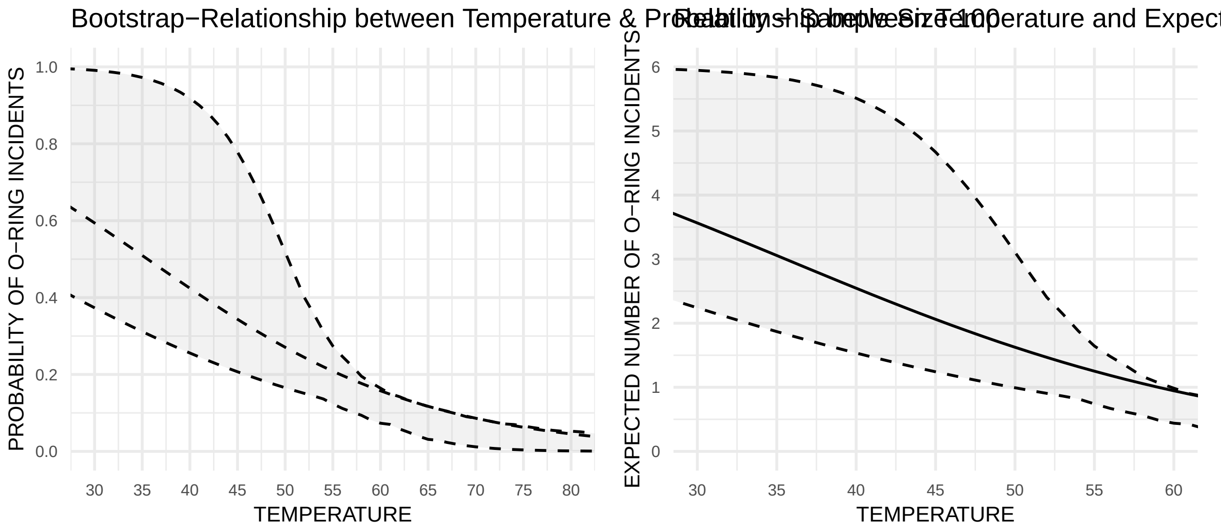
To construct the confidence interval, we utilized those predicted probabilities from the bootstrap resampled datasets. Instead of relying on formula-based methods, we opted for calculating the confidence interval based on quantiles from the bootstrap distribution. This quantile-based approach is more robust for outliers by continuously resampling the data. We chose a 90% confidence interval to represent the range of plausible values. Initially, we tried 1000 bootstrap samples, and then increased the sample size to 30,000 and 50,000. It was observed that the accuracy of the confidence interval improved with larger sample sizes, resulting in narrower intervals.

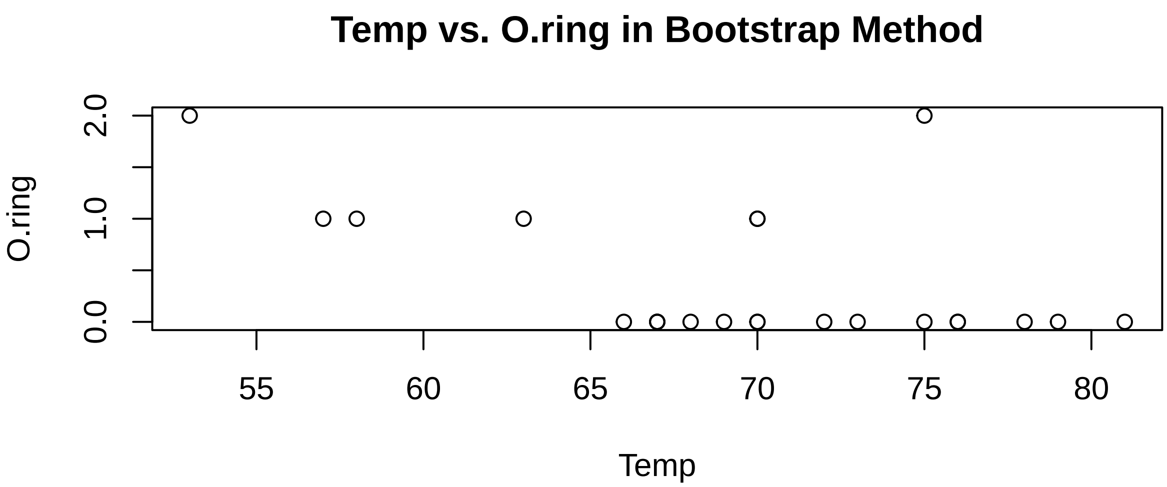
Figure 5.shows these Bootstrap confidence intervals, providing a plot of predicted probabilities with a 90% bootstrap interval that illustrates the probability of incidents in relation to temperature. The upper and lower dots represent 95th and 5th percentiles of the bootstrap distribution, respectively. Notably, the intervals are wider for temperature below 65 F and narrower for temperature above 65 F.

The wider interval for temperature below 65°F can be explained by figure 7, which shows fewer observations shown by small black dots for temperature below 65°F compared to temperatures above 65°F. With fewer observations, the variability in the data is not easily captured resulting in higher standard error. The higher standard error indicated less precision in the estimates, leading to wider confidence intervals for temperature below 65 F. For instance, in fig 6, at a temperature of 30°F, the interval spans the expected number of incidents from 1 to 6 due to high variability and standard error in the data due to less data observations at that point.

An interesting aspect of the Bootstrap method is its ability to account for outliers, as outliers demonstrated in figure 7 at temperatures of 75°F and 55°F. The resampling process of Bootstrap involves randomly selecting subsets of the dataset with replacement, which means that outliers have a chance to be included multiple times in the resampled data. This allows each observation in the original dataset to contribute to the estimation of the confidence interval, effectively accounting for outliers.

In summary, the wider interval for temperature below 65°F indicates higher variability and less precision in the estimate due to few observations. Conversely, the narrower interval for temperature above 65°F suggests lower variability and greater precision. Additionally, the bootstrap method's reampling process addresses outliers by including them in the resampled data, contributing to more robust confidence interval estimates.





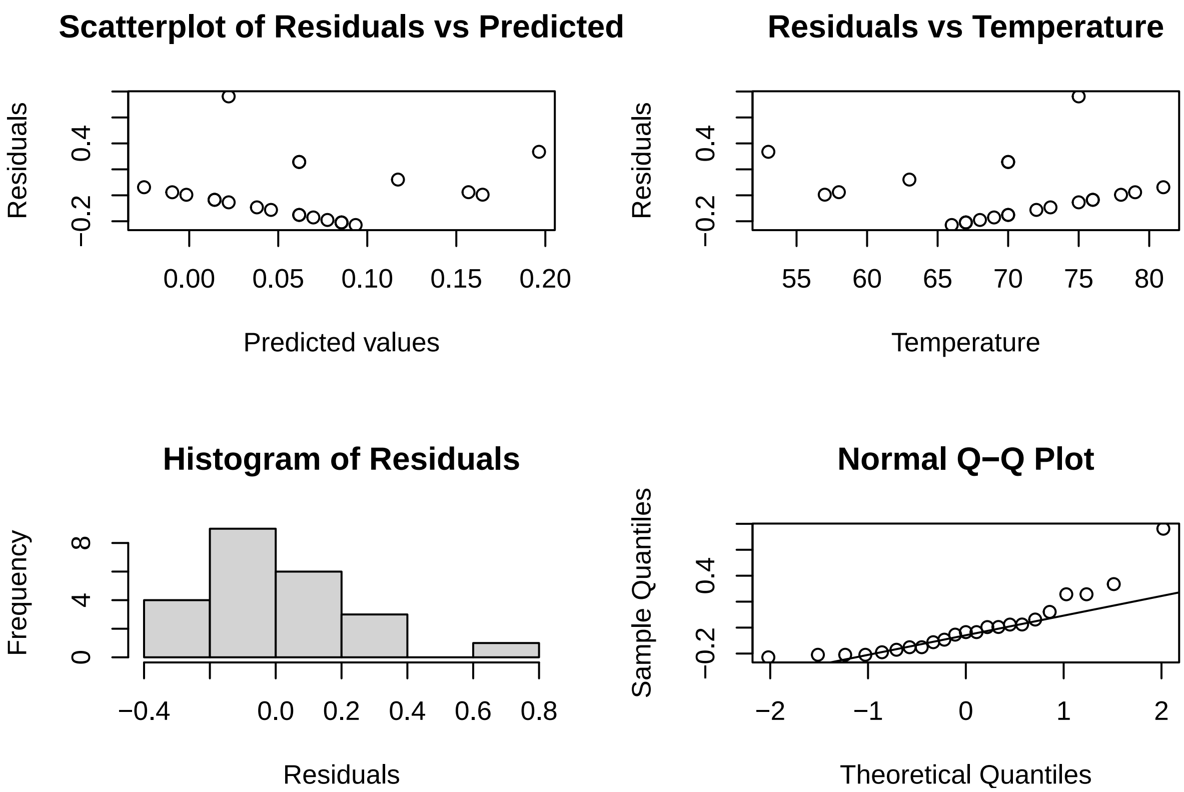
3.4 Alternative Specification

Let's consider an alternative model for analyzing this dataset, namely the linear regression model. The linear regression model is constructed using the response variable "failure rate" and the explanatory variable "temperature". Since the failure observations of O-rings are aggregated at the flight level, we assign equal weights to all the aggregated observations. Upon applying linear regression, we obtain the following estimates. The explanatory variable "Temp" is found to be statistically significant with a coefficient 0.007. This suggests that a one unit increase in temperature is associated with a 0.7% decrease in the failure rate.

While assessing the assumptions of the linear regression model, several aspects need to be considered. Firstly, the normality of residual errors is examined. From the figure 8, histogram of residuals, it can be observed that the distribution is approximately normal. Secondly, homoscedastic assumption is checked by evaluating the pattern of residuals with respect to the predicted values. The residuals do not exhibit any noticeable pattern, indicating homoscedastic. Additionally, the QQ plot shows that the errors are symmetrically distributed around the line, with fewer outliers towards the ends. Furthermore, a Shapiro test for normality yields a significant result, supporting the assumption of normality.

It is important to note that the response variable in this analysis is derived by calculating the ratio of O-ring failure counts to the total number of O-rings. The counts of O-rings range from 1 to 6, representing the six possible outcomes. Therefore, the response variable is a discrete variable rather than a continuous variable. Logistics regression, which is well-suited for modeling probabilities, is more appropriate in this case. Logistics regression handles the probability bounds of 0 and 1 and effectively models the probability associated with a discrete response variable. Moreover logistics regression allows for straightforward interpretation of the response variable in terms of probability or odd ratio.

Considering these factors, the logistics model is deemed more suitable for capturing the relationship between explanatory variables and the discrete response variable. It provides meaningful interpretation in terms of probabilities or odds ratio, allowing for a more comprehensive understanding of data.




```
##
##  Shapiro-Wilk normality test
##
## data:  residuals
## W = 0.8233, p-value = 0.0009267
```

4 Conclusions

The above probabilistic study expands on the risk analysis of the Space Shuttle: Pre-Challenger Prediction of Failure research conducted by Dala, Fowlkes and Hoadley conducted in 1989.

The primary goal of the study is to understand the impact of independent variables temperature and pressure on the dependent variable Number of O-ring failures. Using the Logistic Regression model we are able to conclude that the independent variable temperature is statistically significant in impacting the No of O-ring failures.

We observe that the temperature has statistical significance on the No of O-ring failures. We have also calculated the quadratic temperature term and observe that, there is not sufficient evidence suggesting that adding the quadratic temperature term improves the logistic model. We have also observed that the variable pressure has no statistical significance on the number of O-ring failures.

In addition, we have tested with Wald Confidence interval and log likelihood ratio and the results are statistically significant, that also concluded that lower the temperature higher the probability of the number of 0-ring failures. When the temperature is 31°F, the estimated probability of O-ring failure is 0.8178 with the confidence interval between 0.1596 and 0.9907. The estimated number of O-ring failures at 31°F is 4.9066 with the confidence interval between 0.9576 and 5.9439. In order to draw inference from the model, we need to assume the model is built off a representative data set including records with both high and low temperatures, high and low pressure levels. Note that we are extrapolating our model inference outside of the known temperature range which could lead to error.

As part of the Bootstrap Confidence Intervals we are simulating the sampling (with replacement because of small data size) process over iterations and performed model fitting using the sample, generated in each iteration and used it for prediction and calculation of confidence interval for different temperature values. We are running the simulation 50,000 times to generate a stable distribution. The confidence interval varies depending on the temperature value. At lower temperature, the confidence interval tends to get wider as it moves further away from the center of the data, resulting in the uncertainty of the model predictions.

Finally, we conclude that the launch temperature of 31°F had substantial risk to the Challenger Space shuttle launch, and determine that the shuttle should have been grounded until the risk was mitigated.