W271 Group Lab 1

Investigating the 1986 Space Shuttle Challenger Accident

Jinsoo Chung, Carolyn Dunlap, Ayda Nayeb Nazar, Jocelyn Thai

Contents

Abstract

While the 1986 accident of the space shuttle *Challenger* was tragic, a risk analysis that followed, conducted by Dalal *et al.* (1989), demonstrated the importance of using statistical modelling to predict and risk-manage such high-impact ventures in the future. In our own analysis, we explore the original *Challenger* dataset and regression models that Dalal *et al.* uses to assess the temperature effect on the O-rings used aboard the spacecraft. We determine various confidence intervals for the probability of O-ring failure and discuss the impact of using alternative models. Based on our results, we found a much higher estimated probability for O-ring failure at lower temperatures, indicating that more thorough statistical analyses of even smaller data sets may be extremely vital for risk management of space launches in the future.

1 Introduction

1.1 Research question

Our research question is to investigate the statistical relationship between primary O-ring failure and joint temperature and/or leak-check pressure using both logistic and linear regression models.

2 Data

2.1 Description

The original data set used by Dalal *et al.* for analysis is comprised of 23 observations from flights prior to the *Challenger*'s launch with 5 variables: the identifying flight number (1 through 23), the joint temperature at the time of launch (measured in degrees Fahrenheit), the leak-check pressure used to test O-rings post-assembly (measured in pounds per square inch, psi), the number of primary O-rings that failed, and the total number of primary O-rings surrounding the solid rocket motors (n=6 O-rings for all observed flights). The data was collected after the rocket motors were recovered and inspected after each launch; there were 24 flights launched prior to the *Challenger*, of which 23 had successfully recovered rocket motors and thus data to use in this analysis.

From this data, there are three main variables in question, the temperature, pressure, and number of failed O-rings. The dependent variable, or outcome, is the count of O-ring failure, and the potential explanatory variables are temperature and leak-check pressure. For each explanatory variable, the joint temperature at the time of launch was stochastic and dependent on the temperature on that launch day. The leak-check pressure used for earlier flights was originally set to 50 psi, and increased to 100 psi and finally 200 psi with later flights.

The number of failed O-rings can be considered as either a binary or binomial response variable. As a binomial variable, each O-ring failure is counted out of 6 total trials, given that there are 6 primary O-rings used to seal the solid rocket motors in each launch. Alternatively, O-ring failure can be converted to a binary response, where at least one O-ring failure would be considered as failed and no failure considered as not failed.

Dalal *et al* uses logisitic regression to estimate the probability of O-ring failure, which requires that all data points are independent and identically distributed. In this case, for a binomial response, each of the six O-rings must fail independently. This is a reasonable assumption because each primary O-ring is responsible for sealing different joints in the solid rocket motors, so the failure of one primary O-ring should not impact the failure of another primary O-ring. In the case that O-ring failure is considered a binary response, each O-ring success per flight must be independent from every O-ring success or failure other flight. This is also a reasonable assumption, because new O-rings are used for each flight so the success or failure of O-rings in a previous flight should not influence the success or failure on a future flight.

These indepedence assumptions are necessary; if the O-rings are not independent, the failure of another O-ring will be dependent on the failure of an O-ring, and thus both probability models will have biases. If such behavior occurs, our model will not reflect the behavior of the population.

2.2 Key Features

There were no missing values in the data set. For all flights prior to *Challenger*, there were 6 primary O-rings per spacecraft and either 0, 1, or 2 O-rings failed per flight(counts for each displayed in Table 1). There is a skewed underlying population, with a majority of launches (70%) having no primary O-ring failure. O-rings were tested with either 50, 100, or 200 psi (Table 2); in most launches, O-rings were either tested at 50 or 200 psi (26% and 65% of the data respectively). Temperatures from the flights varied from 53 to 81 degrees F, with a median temperature of 70 degrees, as summarized in Figure 1.

Table 1:

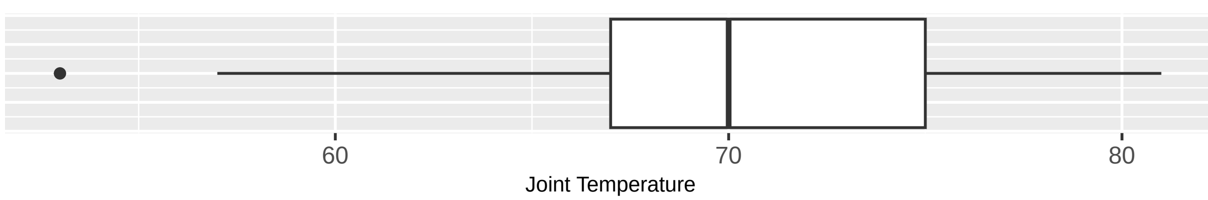
# Failed O-Rings	Count	Proportion
0	16	0.70
1	5	0.22
2	2	0.09

Table 2:

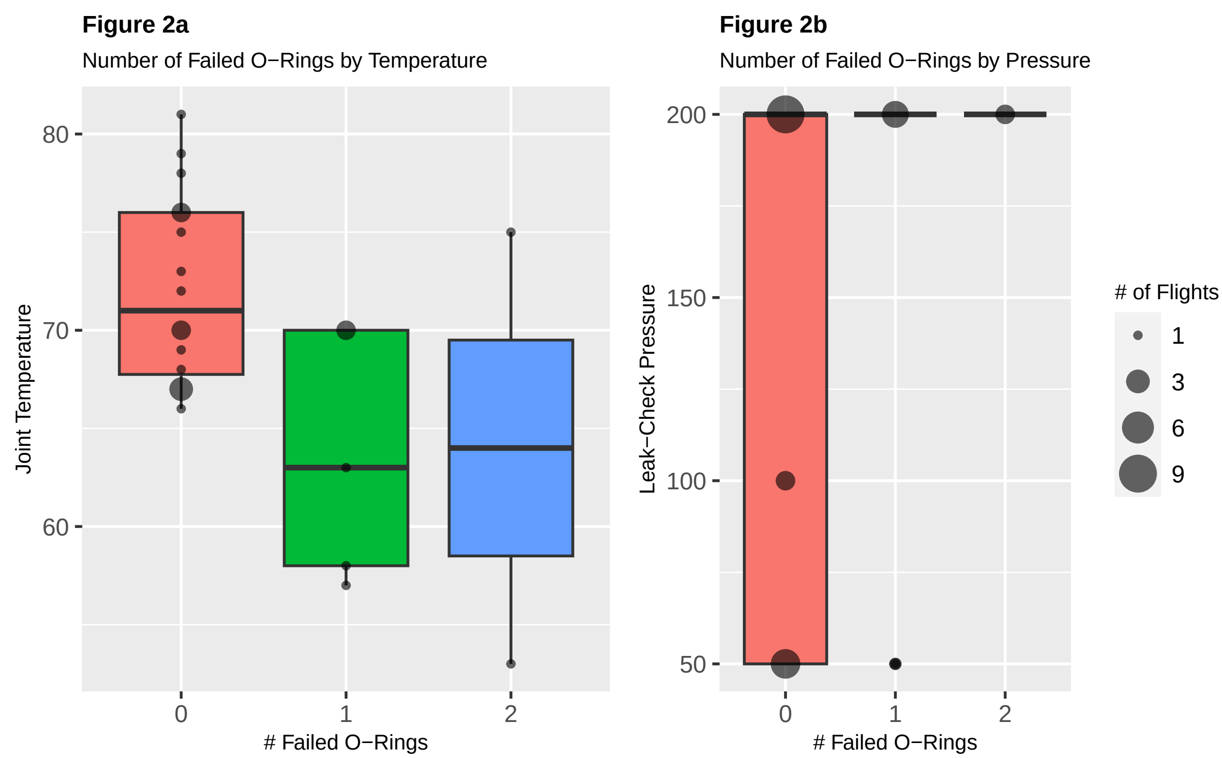
Leak-Check Pressure (psi)	Count	Proportion
50	6	0.26
100	2	0.09
200	15	0.65

Figure 1

Joint Temperature per Flight



Because we are interested in assessing reasons behind O-ring failure, we conducted bivariate analyses of flight temperature or pressure versus the number of failed O-rings, respectively (Figure 2a and 2b). As shown in Figure 2a, there does appear to be a temperature effect on O-ring failure, where instances of O-ring failure are more prevalent at lower temperature. Conversely, there does not appear to be any strong relationship between leak-check pressure and O-ring failure (Figure 2b).



3 Analysis

3.1 Reproducing Previous Analysis

Two logistic models are explored by Dalal *et al* - binomial and binary. We explored both models.

3.1.1 Binomial Model (Model 1):

```
##
## Call:
## glm(formula = O.ring.fail.prob ~ Temp + Pressure, family = binomial(link = "logit"),
##     data = challengerv1)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.42299  -0.26267  -0.21671  -0.06634   0.95601
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.520195   8.540918   0.295    0.768
## Temp        -0.098297   0.109959  -0.894    0.371
## Pressure     0.008484   0.018806   0.451    0.652
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4.0384  on 22  degrees of freedom
```

```
## Residual deviance: 2.7576  on 20  degrees of freedom
## AIC: 9.2286
##
## Number of Fisher Scoring iterations: 6
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: O.ring.fail.prob
##          LR Chisq Df Pr(>Chisq)
## Temp      0.86397  1     0.3526
## Pressure  0.25678  1     0.6123
```

As seen from the p-values of the coefficients using Anova, both temperature and pressure (p-val = 0.3526309 and 0.6123437 respectively) in the binomial regression model are not significant. This means that the coefficients are not significantly different from zero.

On a more practical understanding of Model 1, temperature seems to have more of an effect than pressure on the probability of O-ring failure. The model estimates that for every 1 degree increase in temperature, the probability of O-ring failure will decrease by 1.104 times, assuming pressure remains constant. Conversely, for every 1 psi increase in pressure, the probability of O-ring failure will increase by 1.009 times, assuming temperature remains constant. However, based on the likelihood ratio test, these predicted effects of each variable does not significantly improve the model, as seen from the Chi square probability values of 0.3526 and 0.6123, and so both temperature and pressure do not have a statistically significant effect on the likelihood of O-ring failure using a binomial logistic regression. An alternative model to the binomial regression is to have O-ring failure on each flight be a binary response - did an O-ring fail or not? Because even one O-ring failure can be critical in a space flight, converting the O-ring failure to a binary response is appropriate, and thus we also explored Dalal *et al*'s binary logistic regression model.

3.1.2 Binary Model (Model 2):

```
##
## Call:
## glm(formula = O.ring.fail ~ Temp + Pressure, family = binomial(link = "logit"),
##     data = challengerv2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1993  -0.5778  -0.4247   0.3523   2.1449
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 13.292360   7.663968   1.734   0.0828 .
## Temp        -0.228671   0.109988  -2.079   0.0376 *
## Pressure     0.010400   0.008979   1.158   0.2468
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
##     Null deviance: 28.267  on 22  degrees of freedom
## Residual deviance: 18.782  on 20  degrees of freedom
## AIC: 24.782
##
## Number of Fisher Scoring iterations: 5
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: O.ring.fail
##           LR Chisq Df Pr(>Chisq)
## Temp        7.7542  1   0.005359 **
## Pressure    1.5331  1   0.215648
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Unlike Model 1, temperature is a significant variable in Model 2, in which the coefficient is significantly different from zero (p = 0.005259 by Anova). The coefficient for pressure was not significantly different from zero, suggesting even with a binary logistic regression model, pressure does not have an effect on O-ring failure. The effects of temperature and pressure are similar to Model 1, in which a degree increase in temperature is negatively correlated (beta = -0.2287) with the log odds of failure, and a psi increase in pressure is positively correlated with the log odds of failure. In Model 2, the odds of O-ring failure is 0.80 times as large for every 1 degree increase in temperature (holding all other variables constant), indicating that failure rate decreases as temperature increases.

In the paper, Dalal, Fowlkes, and Hoadley (1989) chose to remove `pressure` from the model based on their likelihood ratio tests. As seen from both Model 1 and Model 2, pressure is not a significant variable. From both Anova tests, pressure was found to not have a significant influence in the models, and the effect of pressure was not significantly different from zero. Using only Model 1, the decision to take out pressure from the model may have been ambiguous, as both temperature and pressure were not significant variables. However, with both models the authors were justified in removing the `pressure` variable from the model, which remained not statistically significant regardless of which model was used. Therefore, for the remainder of this report, we are choosing to remove `pressure` from our model.

3.2 Confidence Intervals

To construct confidence intervals, we considered the simplified binary logistic model $logit(\pi) = \beta_0 + \beta_1 Temp$, where π is the probability of O-ring failure to estimate a logistic regression model. Moreover, we compared the model to a quadratic model $logit(\pi) = \beta_0 + \beta_1 Temp + \beta_2 Temp^2$ to determine whether a quadratic interaction would be significant to the model.

```
##
## Call:
## glm(formula = O.ring.fail.prob ~ Temp + I(Temp^2), family = binomial(link = "logit"),
##     data = challengerv1)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.34423  -0.29551  -0.25303  -0.00545   1.02920
```

```
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 22.126148  58.284482   0.380    0.704
## Temp        -0.650885   1.814484  -0.359    0.720
## I(Temp^2)    0.004141   0.013943   0.297    0.766
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4.0384  on 22  degrees of freedom
## Residual deviance: 2.9319  on 20  degrees of freedom
## AIC: 9.2373
##
## Number of Fisher Scoring iterations: 6
```

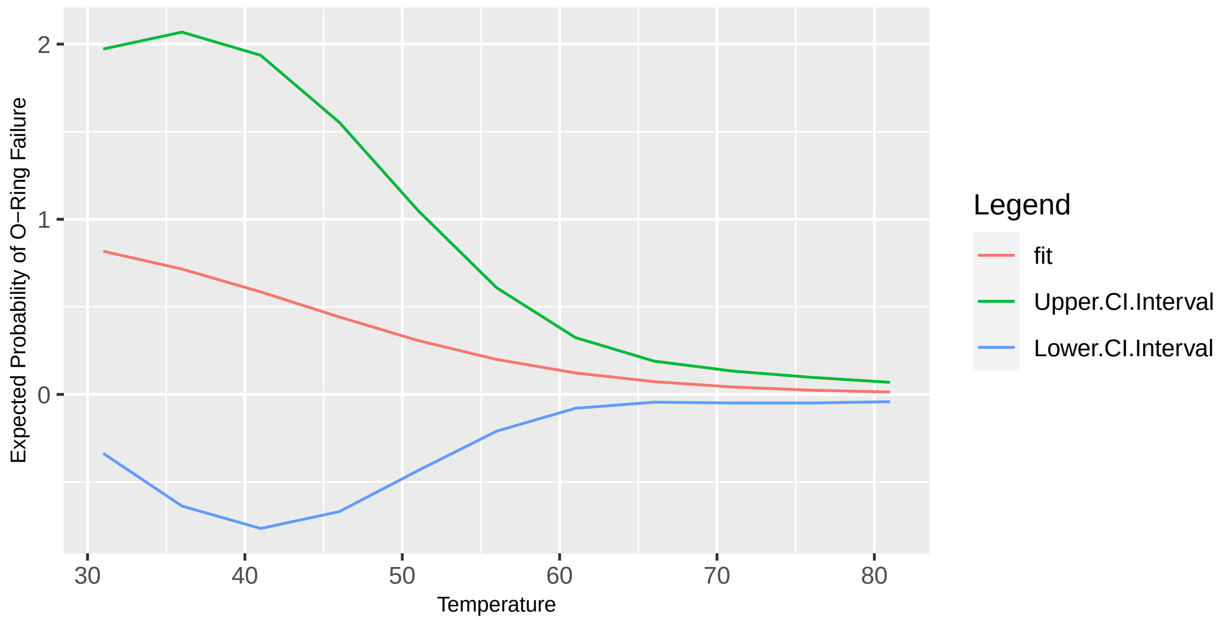
```
## Likelihood ratio test
##
## Model 1: O.ring.fail.prob ~ Temp
## Model 2: O.ring.fail.prob ~ Temp + I(Temp^2)
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   2 -1.6013
## 2   3 -1.6186  1 0.0346     0.8524
```

As seen from the summary of the model with an included quadratic interaction term, the quadratic interaction term is not a significant variable in the model. As seen from the likelihood ratio test, the addition of the quadratic does not add a significant (Pr Chisq = 0.8524) benefit to the model. Thus it's sensible not to add a quadratic term for temperature to our model.

We therefore modeled the probability of failure against temperature using only temperature as the explanatory variable. We considered the temperature ranges from 31 to 81 degrees and included in the plot the 95% Wald confidence interval bands.

Figure 3

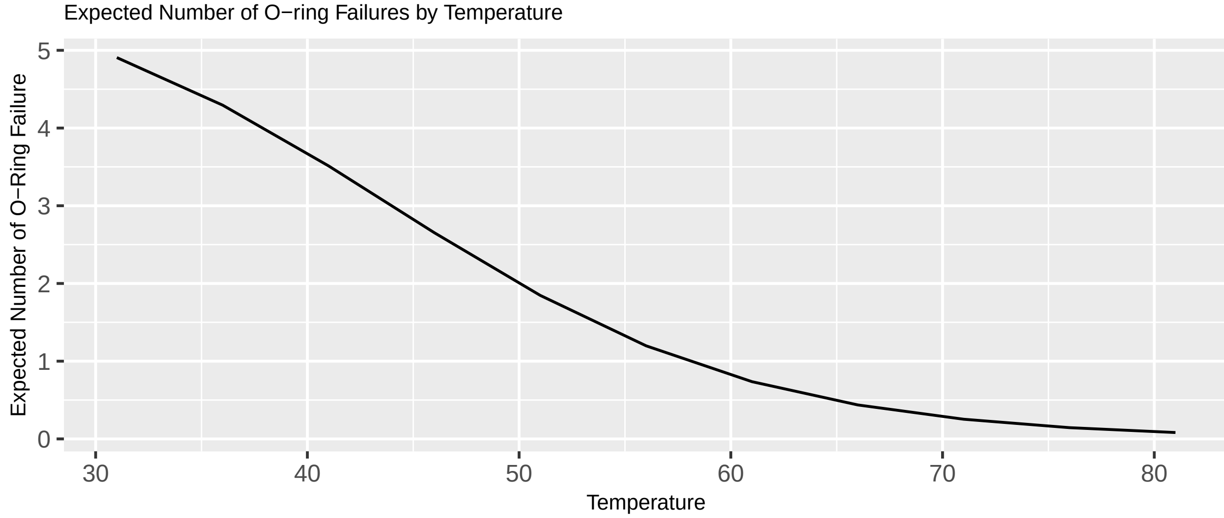
Probability of O−ring Failure by Temperature



As seen from the plot, as temperature increases, the likelihood of O-ring failure decreases. The confidence interval is much greater for lower temperatures, suggesting a higher variability in the prediction model at lower temperatures. One explanation would be due to the distribution of the O-ring failures by temperature. First, there's little to no data in the lower temperature ranges, which adds to greater estimated standard errors since we are extrapolating the model well outside the range of the data (which had a minimum temperature of 53 degrees). Second, most temperature data associated with no failure is clustered in the higher temperature range, which reduces the amount of standard error at higher temperature. In any case, this confidence interval plot suggests that there's much more uncertainty in the predictive model for lower temperature values, using standard confidence intervals.

Figure 4

Expected Number of O−ring Failures by Temperature



```
##         fit Upper.CI.Interval Lower.CI.Interval
## 1 0.8177744          1.972169        -0.3366201
```

The temperature was 31° at launch for the Challenger in 1986. The estimated failure at 31 degrees was 0.8178, with a 95% confidence interval between 1.9722 and -0.3366. This means that there's a high chance that there's around 4 to 5 primary O-rings that failed at that temperature. In order for this estimation to hold, we'd need to assume that the temperature dependent O-ring failure probability follows a binomial logistic probability model, which means that the log odds of failure are linearly related to temperature. This means that there's no major property change in the O-ring at the specified temperature or there's no major separate environmental variable, such as condensation, that may significantly alter the behavior of the O-ring. Because our data also does not contain any data points near the 31° launch of the Challenger, we are relying on the asymptotic properties of the model to determine both the confidence interval and estimated probability of O-ring failure.

3.3 Bootstrap Confidence Intervals

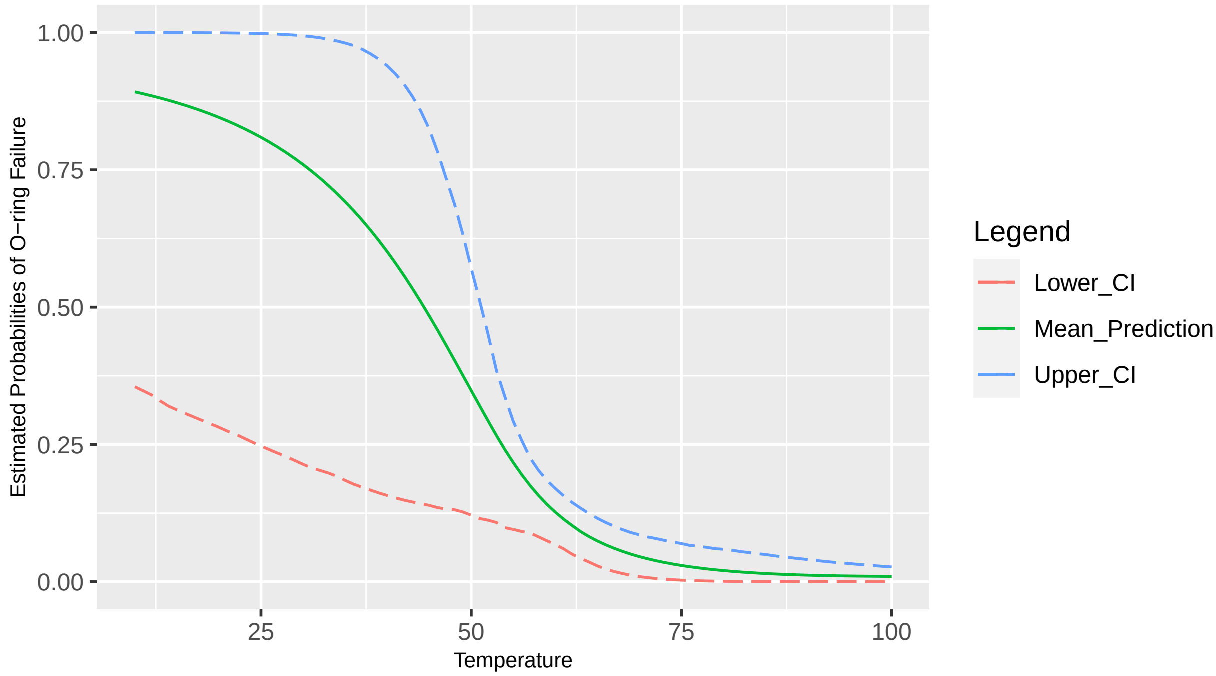
Rather than relying on asymptotic properties, we considered using a parametric bootstrap, as did Dalal, Fowlkes and Hoadley. We simulated a large number of data sets (n = 23 for each) by re-sampling with replacement from data, and examined the 90% confidence intervals of the models.

	Temp	Lower_CI	Upper_CI	Mean_Prediction
5%	10	0.3549713	0.9999489	0.8919930
5%1	11	0.3475757	0.9999350	0.8884817
5%2	12	0.3400535	0.9999175	0.8847622
5%3	13	0.3293547	0.9998954	0.8808201
5%4	14	0.3197527	0.9998673	0.8766401
5%5	15	0.3129608	0.9998317	0.8722055

We performed 1000 simulations and resampled datasets of size 23 by utilized the `sample` function and setting `replace=TRUE` in order to do the sampling with replacement. We then fit a logistic regression model to each re-sampled data set and utilized the mean of the predictions from all the models to get a final estimated probability of O-ring failure at every temperature between 10° and 100° Fahrenheit, with the head of the data shown in the table above. 90% confidence intervals were obtained for each of these temperatures by extracting the 0.05 and 0.95 quantiles of our model predictions, as was done by the authors. The plot below illustrates the predicted probabilities for each temperature.

Figure 5

Estimated Probabilities of O−ring Failure vs. Temperature from Bootstrapping



By using this bootstrap method, we show that while we still have much larger confidence intervals at lower temperatures, for the same reasons as discussed in the previous section, our confidence intervals are now appropriately constrained by the probability limits of 0 and 1, which was previously a large point of concern with our previous confidence intervals.

3.4 Alternative Specification

Since we had been largely working with logistic regression models, we also decided to alternatively assess a linear regression models of O-ring failure against temperature, as described below:

```
##
## Call:
## lm(formula = O.ring.fail.prob ~ Temp, data = challengerv1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.09347 -0.06573 -0.01423  0.01760  0.31118
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.616402   0.203252   3.033  0.00633 **
## Temp        -0.007923   0.002907  -2.725  0.01268 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09624 on 21 degrees of freedom
## Multiple R-squared:  0.2613, Adjusted R-squared:  0.2261
```

```
## F-statistic: 7.426 on 1 and 21 DF,  p-value: 0.01268
##
##  Box-Ljung test
##
## data:  model5$resid
## X-squared = 0.16961, df = 1, p-value = 0.6805
```

The model estimates that for every 1 unit increase in temperature, there is a predicted 0.007923 decrease in the probability of O-ring failure. The p-value for the coefficient on temperature is 0.01268, which is significant at the 95% level.

Based on the Ljung-box test, the p-value for the residuals was 0.6805, meaning that we fail to reject the null hypothesis that there is no autocorrelation. This is in support that we can trust the model.

Additionally, the linear model has the effect of temperature on O-ring failure at a significant level compared to the binary model. Therefore, using the linear model may appear to be appropriate as it passed both the model diagnostics and has a significant effect when only looking at the models themselves. However, it makes more practical sense to use the binary logistic regression because O-ring failure is a binary value, whether failure occurred or not. It is not on a continuous scale where an explanatory effect results in a subsequent decrease or increase of failure.

4 Conclusions

Based on our exploration and assessment of various models (binomial logistic regression, binary logistic regression, and linear regression) to assess the relationship between primary O-ring failure and temperature, we decided that our preferred model is the binary logistic regression. More importantly, while we assessed the binomial regression model the most, a binomial model assumes that each O-ring failure is independent of every other O-ring failure on the spacecraft. While this assumption may be sufficient since each O-ring is operating to seal a different joint, in reality it is highly likely that the failure of a primary O-ring on one joint may compromise the spacecraft in such a way as to affect the performance of an O-ring on other joints on the same craft. The benefit of a binary model is that we no longer need each O-ring failure on one spacecraft to be independent, which in reality may be the more appropriate and conservative assumption, especially considering enormous cost and potential loss of life associated with each space launch. Additionally, the binary logistic regression model did show that temperature had a statistically significant effect on O-ring failure, both using Wald p-values (p= 0.0376) and by Anova (p= 0.005259).

Based on this model, we saw that the odds of at least 1 O-ring failure increased 1.257 times for every 1 degree drop in temperature, which corresponds to an 9.843 times increase in the odds of O-ring failure for every 10 degree drop in temperature. These results indicate that regardless of pressure (when assessing pressures of 50, 100, or 200 psi as was assessed in the data set), there is an estimated 99.9% probability of at least 1 O-ring failure with a 31 degree temperature, as was seen during the *Challenger*.

These results, along with the results of the binomial model discussed at length in the paper, all point to the statistically significant temperature effect on O-ring failure. We demonstrated statistical proof of this effect using data points that all took place prior to *Challenger*'s launch, which suggests that had this level of analysis been considered, this effect might have been noticed and such a tragedy may have been prevented. For the future, we equally demonstrate the capacity for statistical models to play an important role in risk management moving forward and hope that these analyses can be done to help prevent disaster and improve our space programs and initiatives.