

W271 Group Lab 1

Investigating the 1986 Space Shuttle Challenger Accident

Adam Kreitzman, Hailee Schuele, Lee Perkins, Paul Cooper

Contents

1	Introduction	2
1.1	Research question	2
2	Data	2
2.1	Explaining the Variables	2
2.2	Description	3
2.3	Key Features	3
3	Analysis	5
3.1	Reproducing Previous Analysis	5
3.2	Confidence Intervals	7
3.3	Bootstrap Confidence Intervals	8
3.4	Alternative Specification	9
4	Conclusions	10

Abstract

In this report we reconstruct the Dalal et al analysis surrounding the conditions of the 1989 Challenger rocket launch. We use multinomial logistic regression to predict O-ring failure, the reported cause of the launch catastrophe. We learn that temperature has a statistically and practically significant effect on O-ring failure. Like Dalal, we conclude that postponing the Challenger launch to a warmer temperature would have increased the probability of a success.

1 Introduction

In 1986, the Challenger Space Shuttle catastrophically exploded during launch. Subsequently, the United States conducted an analysis of the incident, and the Rogers Commission published a report that concluded that the explosion was caused by faulty O-rings in the rocket motor engine assembly. Additional analysis, performed by Dalal et al, concluded that NASA had enough information from prior rocket launches to know that O-ring failure was likely on a day as cold as the Challenger launch.

1.1 Research question

We are conducting a risk analysis based on data from 23 separate shuttle launches prior to the Challenger. In particular we are investigating whether or not Temperature and Pressure have an appreciable impact on O-ring failures. Our response variable is the number of thermally distressed primary O-rings, which can range from 0 to 6 (since there are six primary O-rings per shuttle). We will be using a binomial distribution to model the number of distressed O-rings, where the number of successes (distressed O-rings) follows a binomial distribution with a fixed number of trials (6) and a probability of success (p).

Our two potential explanatory variables are Temperature, which is measured in degrees Fahrenheit, and Pressure, whose unit of measurement is pounds per square inch (psi).

We hope to determine if there is a significant correlation between one or both of these variables to be able to make safety recommendations for limiting catastrophic O-ring failure in shuttle launches.

2 Data

2.1 Explaining the Variables

Temperature

The temperature variable is the temperature at the moment when the O-rings were examined for failure. The temperature ranges from 53 to 81 degrees Fahrenheit in the dataset.

Pressure

Pressure is a categorical variable, indicating the leak test pressure, in psi, of the O-rings for a given flight. This increased over time, as pressure thresholds increased over the history of NASA launches.

O.Rings

O-rings are an important structural component of rocket launches that seal the field joints of the solid rocket boosters during the launch into orbit. The O-rings data from the 23 past launches

is the count of how many O-rings experienced thermal distress on that flight. The data ranges between 0 and 2 failures per launch, with a theoretically possible maximum value of 6.

Independence

Independence is a necessary assumption of a binomial model, as a lack of independence could imply that the probability of O-ring failure is not consistent for each trial. This may be difficult to justify in this context. Each launch contains 6 field joints, and the failure of a joint may reasonably be expected to contribute to the failure of other joints on that same launch. Dalal et al were able to alleviate this concern by running a subsequent analysis, where they measured whether at least 1 O-ring failed on a given launch, which provided similar results and justified the assumption of independence.

2.2 Description

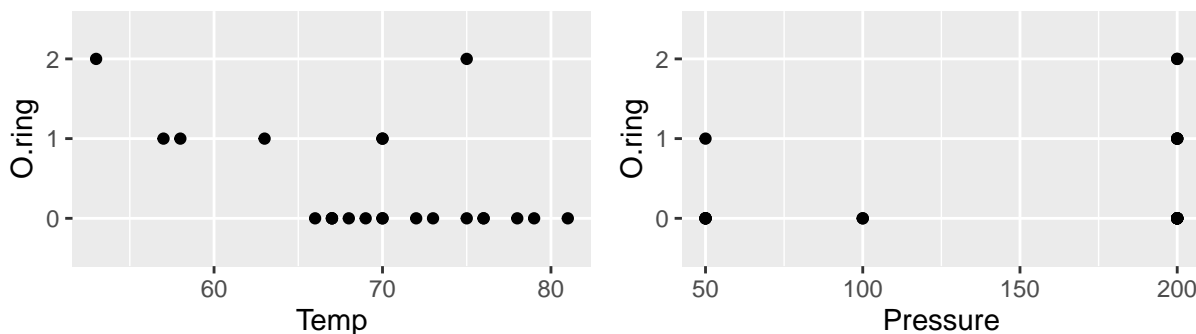
The Challenger dataset consists of 23 observations from shuttle flights conducted prior to the Challenger launch. Observations of O-ring failures were gathered by collecting the rocket motors from the ocean after the launch. Temperature was recorded on the day of the launch, and pressure was measured as the leak test pressure of the O-rings during their assembly. Our population of interest is O-rings at any reasonable pressure and temperature from the 4-part rocket motor assembly used by NASA during the Challenger launch.

The variables of interest are described in the previous section. None of these variables have missing values and there are no outliers. While all are of numeric type, it's worth noting Temperature is continuous, Pressure is only recorded as 3 values (50, 100, 200), and O-rings only has 3 values (0, 1, 2). All observations were retained for the analysis.

2.3 Key Features

Our EDA was mainly concentrated on looking at how the two explanatory variables (Temperature and Pressure) interacted with the response variable (O.Rings).

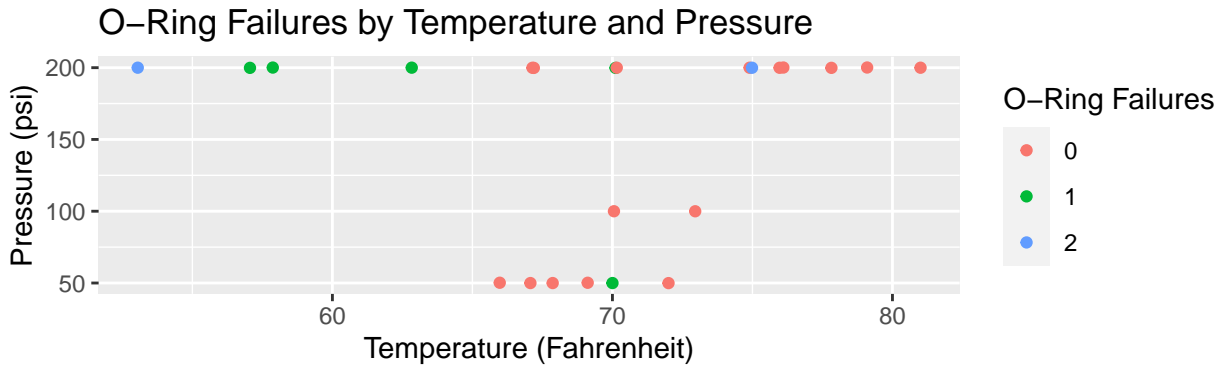
The first chart that we created was a scatterplot of Temperature vs. O.Rings, seen below. For the most part, O-ring failures happen at lower temperatures. There's one instance of an O-ring failure at 70 degrees and 2 O-ring failures at 75 degrees. Otherwise, we see every instance below 65 degrees resulting in at least 1 O-ring failure.



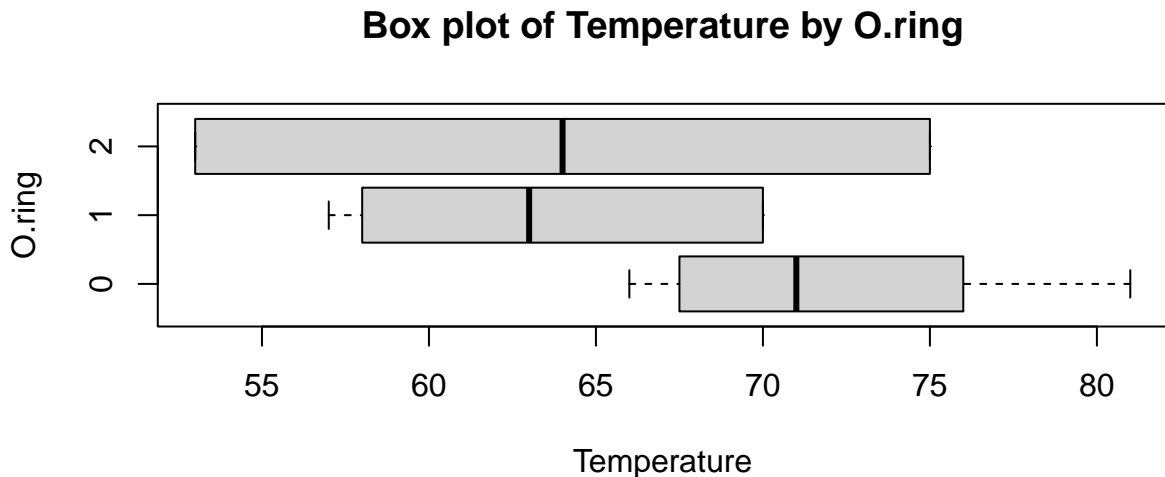
The for the graph on the right we created was a scatterplot of Pressure vs. O.Ring. This scatterplot does not make pressure seem to be a relevant factor, as we see both low pressure and high pressure have instances of O-ring failure, though it is not definitive because there are many points of overlap. It's interesting to see that at a Pressure of 100, there are no instances of O-ring failure, so there

is the possibility that only extremes on either side could influence failure, but more inspection is needed.

This next scatter plot combines all three variables of interest. Again, we can see more O-ring failures clustered at lower temperatures. Failures seem to be more common at higher pressures as well, although this could be a trick of the eye as there are many more data points at higher pressure than lower pressure. Note some space is added to show points that overlap.

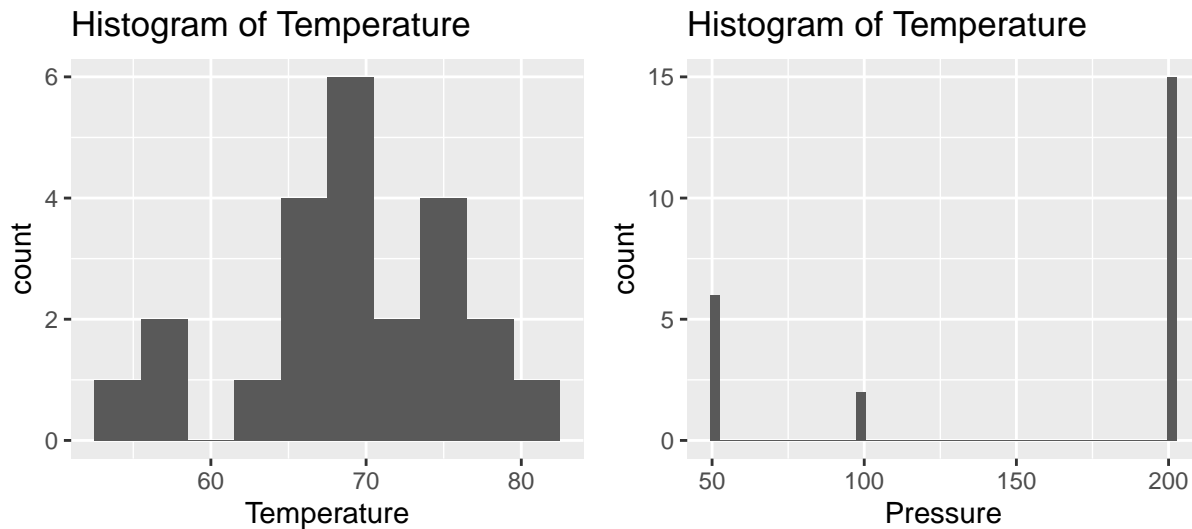


In order to more closely visualize the difference in Temperature by class, we created comparative boxplots that were sorted into the number of O-ring failures for Temperature, which is below:



We can see a very clear upward shift in Temperature for both IQR and Median for the observations without O-ring failures versus the observations with either 1 or 2 failures.

Finally, we wanted to know about the distribution of our data for both Temperature and Pressure so that we know which ranges of variables will have the most robust analysis, so we generated the two histograms below:



For Temperature, we see that the vast majority of our observations lie between 65 degrees and 80 degrees. Our Pressure histogram is a bit funky as the data is only measured at 3 levels, but we can see the majority of measurements are at 200 PSI, with the second most being at 50 PSI, and the fewest at 100 PSI. We would expect, based on this, our analysis to be the most robust between 65 and 80 degrees and at 200 PSI.

3 Analysis

3.1 Reproducing Previous Analysis

3.1.1 Logistic Regression Model

The model below imitates the logistic regression model from Dalal et al. Both explanatory variables are included as linear terms. The outcome variable specifies the count of success (O.ring) and the count of failures (6 - O.ring) to fit the binomial response requirements. The 6 here represents the total number of trials per launch, i.e. the 6 O-rings in each shuttle.

It's worth taking a moment to acknowledge the flipped logic - The “successes” for the binomial model are defined as the number of O-ring failures. Conversely, the model “failures” are the number of O-rings that did not fail.

```
# Logistic regression from the paper
mod_full <- glm(cbind(O.ring, 6 - O.ring) ~ Temp + Pressure,
               data = challenger, family = binomial(link = "logit"))
summary(mod_full)
```

```
##
## Call:
## glm(formula = cbind(O.ring, 6 - O.ring) ~ Temp + Pressure, family = binomial(link = "logit"),
##      data = challenger)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.520195   3.486784   0.723   0.4698
```

```
## Temp          -0.098297    0.044890   -2.190    0.0285 *
## Pressure       0.008484    0.007677    1.105    0.2691
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 24.230  on 22  degrees of freedom
## Residual deviance: 16.546  on 20  degrees of freedom
## AIC: 36.106
##
## Number of Fisher Scoring iterations: 5
```

The summary results here show a coefficient of -0.098 for temperature and 0.008 for pressure. The negative relationship between temperature and O-ring suggests that for each 10 degree increase in temperature, the estimated log odds of success (i.e. O-ring failure) decrease by ~ 1 ($-0.098 * 10$), all other variables are held constant. In other words, the estimated odds of O-ring failure change by 2.67 times for a 10 degree decrease in temperature. On the other hand, for each 10 psi increase in pressure, the estimated log odds of O-ring failure increases by 0.08. This translates to a 1.08 change in odds for a 10 psi increase in pressure.

The effect size for temperature is practically significant, while the effect of pressure is less so. That said, only the coefficient for temperature is significant.

3.1.2 Likelihood Ratio Tests

To further explore the importances of our explanatory variables, we conducted a likelihood ratio test. The likelihood ratio test uses a Type II (i.e. partial) test, which, for each variable, compares a model with the variable to a model without the variable. The results below affirm our prior conclusion: only temperature is significant to the model.

```
Anova(mod_full, test = "LR")

## Analysis of Deviance Table (Type II tests)
##
## Response: cbind(O.ring, 6 - O.ring)
##      LR Chisq Df Pr(>Chisq)
## Temp      5.1838  1    0.0228 *
## Pressure  1.5407  1    0.2145
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3.1.3 Pressure Debate

P-values from both the model summary and the likelihood ratio test suggest that pressure is not an important variable to the model. In general, it's advised to use the most parsimonious model. However, given the small sample size and that there are only two explanatory variables in the data, we argue that pressure does have some value. At the very least, one could argue there isn't enough data to support dropping pressure without a second thought. By opting to keep pressure, the effect of temperature is lessened (the temperature coefficient shrinks from -0.116 to -0.098). This means

that any conclusions drawn from a temperature-only model are likely more dramatic than those drawn from a temperature + pressure model.

3.2 Confidence Intervals

3.2.1 Estimating Temperature Only Models

The models below use temperature as the only explanatory variable. The first (`mod_temp`) uses linear temperature. The second (`mod_temp_quad`) uses linear and quadratic temperature. The final `anova()` test compares the two temperature models to determine the importance of quadratic temperature.

```
# Estimate model with only temperature
mod_temp <- glm(cbind(O.ring, 6 - O.ring) ~ Temp,
               data = challenger, family = binomial(link = "logit"))

# Estimate model with temperature and quadratic term
mod_temp_quad <- glm(cbind(O.ring, 6 - O.ring) ~ Temp + I(Temp^2),
                    data = challenger, family = binomial(link = "logit"))

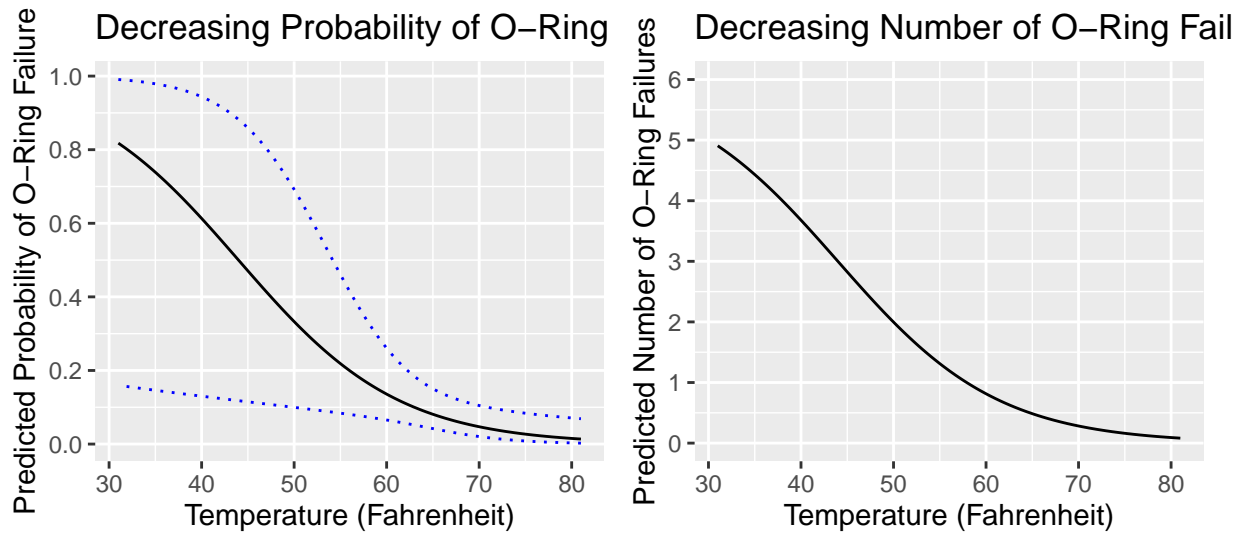
# Determine if quadratic term is important to the model
anova(mod_temp, mod_temp_quad, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: cbind(O.ring, 6 - O.ring) ~ Temp
## Model 2: cbind(O.ring, 6 - O.ring) ~ 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
```

The high p-value in the results of the `anova()` test suggest that the quadratic term is not important to the model. In other words, there is a more linear relationship between temperature and O-ring failure.

3.2.2 Plots

The first graph below shows the probability of O-ring failure across a range of temperatures. The blue bands represent the 95% Wald confidence interval. It's evident that the predicted probability of failure decreases significantly as temperature goes up. The confidence bands are also much wider at lower temperatures. This could be because there are less observations in the data at lower temperatures. The lowest recorded temperature is actually only 53 degrees, so everything below that is a generalization of the model from the rest of the data.



The graph on the right shows the expected number of O-ring failures across temperatures. Again we see the downward trend as temperature increases. It's worth noting that almost 5 O-ring failures are expected at 31 degrees. We'll take a closer look in the next section.

3.2.3 Probability of Failure

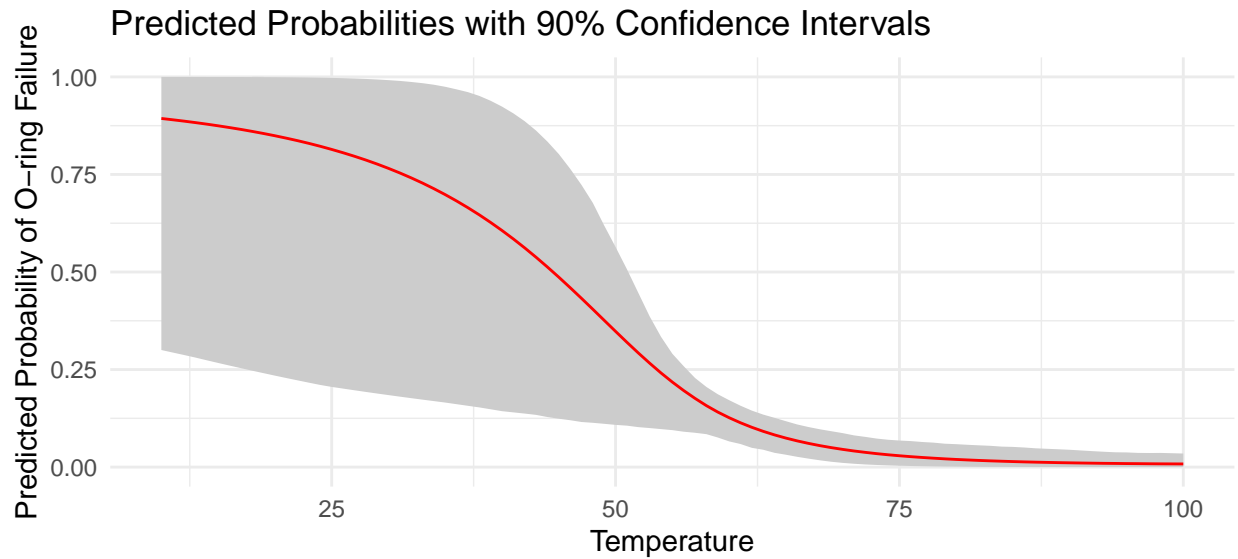
The predicted probability of an O-ring failure at 31 degrees, the temperature during the 1986 Challenger launch, is 0.818. Because of the small number of observations in the data, we opted to use the 95% profile likelihood confidence interval instead of Wald. This resulted in a confidence interval of (0.142, 0.99). This is a very wide range of confidence, which coincides with the “Decreasing Probability of O-Ring Failure with Increased Temperature” graph.

3.2.4 Model Assumptions

1. Two possible outcomes: Each O-ring either fails or doesn't, so this assumption is met.
2. Constant probability of failure: The probability of failure is constructed to depend on temperature; we have reason to believe that this probability function is constant across trials.
3. Independent & identical trials - As discussed above, independence is tricky to establish when O-rings on the same launch may be expected to have some impact on each other. By running a separate analysis on the probability of at least one failure – a scenario that avoids the same-flight dependencies – and getting similar results, we can see that this assumption isn't detrimental to our results.

3.3 Bootstrap Confidence Intervals

For bootstrapping, we elected to do 1,000 samples, as we felt that was significantly large enough to garner the result we were looking for. Also, to keep things consistent, we made sure to sample with replacement to get accurate results. The plot for the confidence interval is below:



As we can see, bootstrapping made the low-temperature predictions more conservative (note that the dimensions of this graph go from 10-100, so it's a bit different from above.. Additionally, we see that the confidence interval is a bit narrower, though that can also be attributed to the fact that we used a 90% confidence interval rather than 95%. At the same time, the sampling with replacement does not really help us much at lower temperatures, as this is still missing from our model, which is why we see a very large interval below 50 degrees.

Our bootstrap method was effectively to sample our dataset with replacement 23 times with 1,000 iterations. This gave us 1,000 datasets of identical $n=23$, generated by sampling our original dataset with replacement. We then generated a model of each bootstrapped dataset and made predictions ranging from $\text{temp} = (10, 100)$. Then, we generated intervals based on the predictions for each dataset.

For our specific variable of interest, we decided to look at 70 degrees, as it was a common value in our dataset and we believed it was an interesting one to look at since it is a likely temperature to see in the real world.

Based on the results, we obtain a predicted value and 90% confidence interval of:

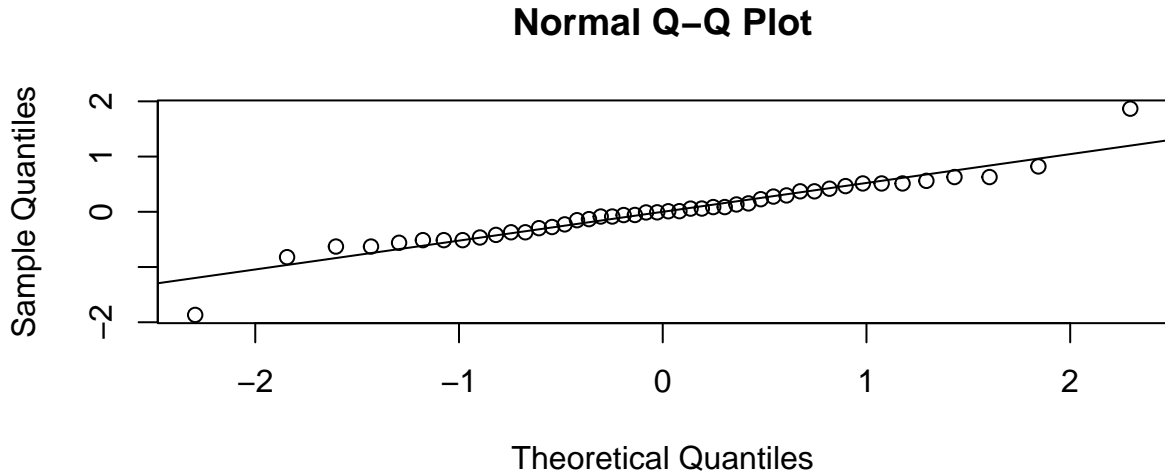
Predicted value: 0.045365556

90% Confidence interval: (.01201458, 0.08518981)

Interpreting these results tells us that our predicted probability of an individual O-ring failure at 70 degrees is roughly .045 (or 4.5%), and we can be 90% confidence that the true probability of an individual O-ring failure at 70 degrees is between roughly .012 (or 1.2%) and roughly .085 (or 8.5%).

3.4 Alternative Specification

We generated a linear model based on Temperature and then proceeded to run tests to check the linearity of the data. The results were clear from examining the pattern of residuals to see if they were randomly distributed. To do so, we generated a QQ plot (below) in order to check for homoscedasticity.



We can see that the residuals are not randomly distributed in the QQ plot. This tells us that a linear model is likely not a good fit for the data. We see that at a lower quantile, the residuals tend to be highly positive, but at higher quantiles the residuals tend to be highly negative. In the middle of the data, it tends to oscillate positive and negative with smaller absolute residuals. Having these kinds of patterns in the QQ plot of linear residuals is not what you will see in a dataset that should be modeled linearly.

4 Conclusions

The logistic model including both temperature and pressure, `mod_full`, is the preferred model to help make safety recommendations for limiting catastrophic O-ring failures in shuttle launches. What makes `mod_full` the preferred model is the negative relationship between the temperature and O-rings. This suggests that for each 10 degree decrease in temperature, the estimated odds of O-ring failure changes by 2.67, all other variables held constant. If we test the probability of failure for the Challenger's conditions with a Temperature of 31 and a pressure of 200, there was a 76.31% chance of failure. Since the dataset was small, the resulting confidence interval has a wide range making it difficult to accurately predict the chance failure. However, we did learn that temperature has a statistically significant effect on the O-ring's chance of failure. This leads us to believe that if the Challenger launched when the temperature was warmer, it would have had a lower chance of failure.