

W271 Group Lab 1

Investigating the 1986 Space Shuttle Challenger Accident

Ray Cao, Ken Trinh, Lisa Wu, Sophie Yeh

Contents

1	Introduction	2
1.1	Research question	2
2	Data (20 points)	2
2.1	Description	2
2.2	Key Features	3
3	Analysis	4
3.1	Reproducing Previous Analysis (10 points)	4
3.2	Confidence Intervals (20 points)	5
3.3	Bootstrap Confidence Intervals (30 points)	7
3.4	Alternative Specification (10 points)	9
4	Conclusions (10 points)	9

Abstract

In this study, we are investigating the effect of physical condition such as temperature and pressure have on the odds of O Rings failure on a space shuttle.

1 Introduction

1.1 Research question

How do temperature and pressure affect the odds of o-ring failures on a shuttle launch?

2 Data (20 points)

We conducted a thorough EDA analysis to gather information of the two potential explanatory variables (temperature and pressure) and the response variable (O.ring). The plots below provide distribution of the three variables and the scatter plot of Temperature vs O.ring.

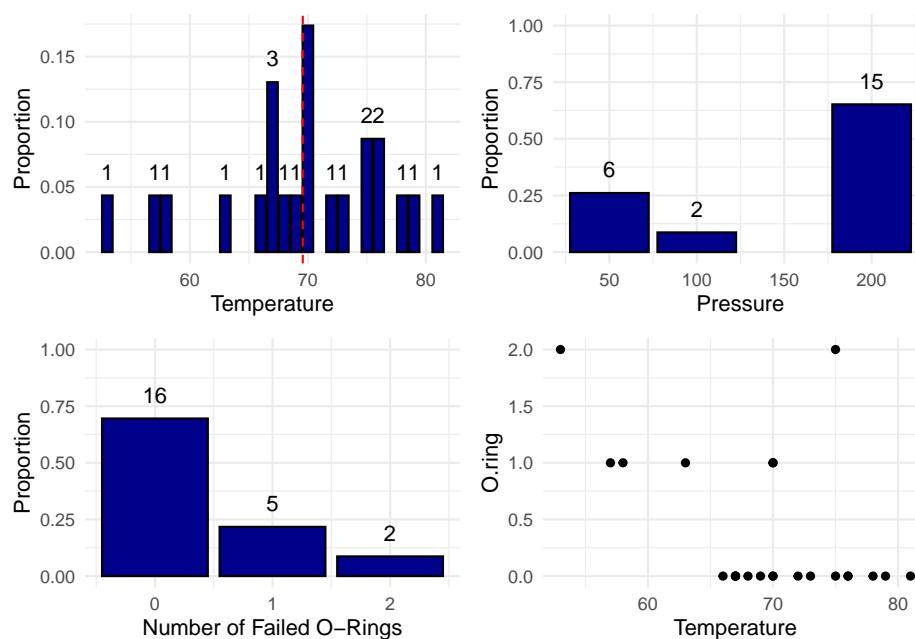


Figure 1: Distribution and Scatter Plots of Key Variables

2.1 Description

This data set was sourced from 23 space shuttle launches before the Challenger launch. It contains 5 columns: Flight, Temp, Pressure, O.ring, Number. Flight is a unique identifier representing each launch. Number represents the total number (6) of O-rings in each shuttle. O.ring is the response variable and denotes the number of O-rings failed at each launch. Lastly, Temp and Pressure (the two potential explanatory variables) denote the numerical values measured at each launch.

There are no missing data for any of the five columns in the data set. Temperature has 16 distinct values, ranging from 53 to 81 with a mean of 69.57. Pressure has 3 distinct values (50, 100, 200) with a mean of 152. Both temperature and pressure records are full integer value. For the response variable O.ring, 16 out of 23 launches have 0 failed O-ring, 5 observations have 1 failed O-ring and 2 observations have 2 failed O-rings.

In the original study, the researchers treated each of the 6 O-rings as independent for each launch. In other words, the failure of one O-ring has no impact on any other O-rings. This assumption is necessary to satisfy the independence assumption of the binomial model. The binomial model also assumes that each O-ring has equal probability of being damaged by either temperature or pressure. By assuming independence, the researcher did not account for the fact that a single failed O-ring could cause the other o-ring or the entire launch to fail. While careful evaluation of the potential dependency between each O-ring is necessary, we will assume independence in this analysis. A binary logistic regression could alleviate the independence concern.

2.2 Key Features

From the plots in Figure 2, we observed that O.ring failures (1 or 2) tend to occur at lower temperature. Higher temperature launches tend to have fewer O-ring failures. In term of pressure, we did not see clear evidence of the impact of pressure on O-ring failures, as zero O-ring failures are evenly distributed across all pressure points. 5 incidents of 1 failed O-ring happened at both low and high pressure and 2 incidents of 2 failed O-rings occurred at high pressure.

We will use both temperature and pressure as explanatory variable to evaluate their effect on the odds of O-ring failures. This will help us answer the research question whether temperature and pressure affect the odds of O-ring failures on a shuttle launch.

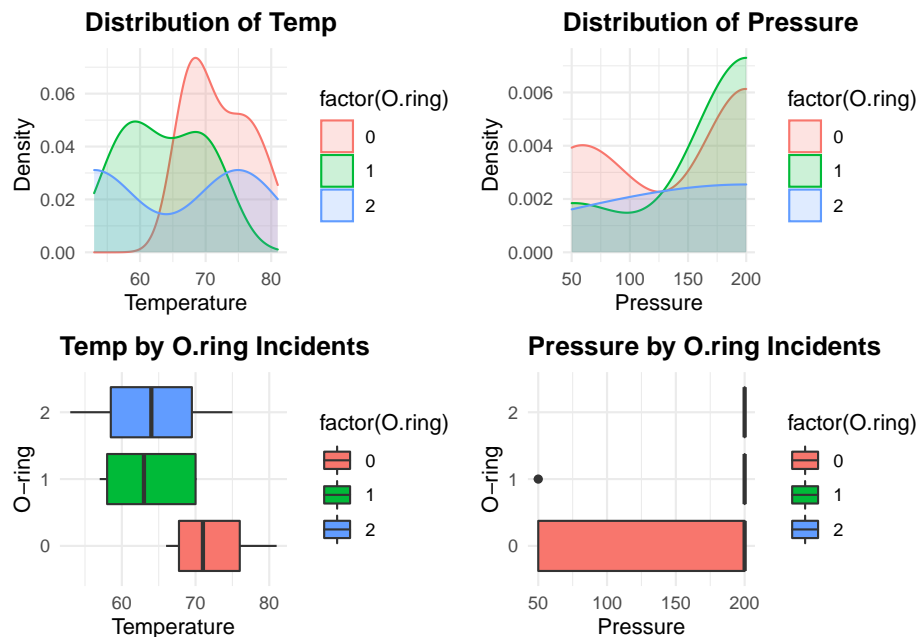


Figure 2: O-ring Failure Distribution Plots for Temperature and Pressure

3 Analysis

3.1 Reproducing Previous Analysis (10 points)

Our baseline model is a binomial logistic regression model that replicates what the authors presented in their original report. See R codes and the model summary table below.

```
mod1 <- glm(formula = cbind(O.ring, Number - O.ring) ~ Temp + Pressure,  
            family = binomial(link = "logit"), data = challenger)
```

Table 1: The Estimated Relationship Between O-Rings Failure and Temperature/Pressure

Output Variable: Log Odds of O-Rings Failure	
Temperature	-0.098* p = 0.029
Pressure	0.008 p = 0.270
(Intercept)	2.520 p = 0.470
Observations	23
Log Likelihood	-15.053
Akaike Inf. Crit.	36.106
Note:	*p<0.05; **p<0.01; ***p<0.001

```
#Impact on odds of O-ring failure for a 10-unit increase of temperature or pressure  
OR1 <- round(exp(10 * mod1$coefficients), 3)[2]  
OR2 <- round(exp(10 * mod1$coefficients), 3)[3]
```

In this baseline model, we perform a logistic regression to estimate the odds of O-ring failure using both temperature and pressure as variables

The estimated model is $y = 2.520 - 0.098 * Temp + 0.008 * Pressure$.

Using a Wald test, the model returns a p-value of 0.029 for temperature and 0.270 for pressure. Since the p-value for Temperature is below 0.05 (the 95% confidence level), Temperature is statistically significant. The p-value for Pressure is above 0.05 and is not statistically significant at the 95% confidence level.

Additionally, temperature have a coefficient of -0.098, indicating that an increase in temperature decreases the log odds of O-ring failures. While pressure is not statistically significant, the positive coefficient suggested that an increase in pressure increases the log odds of O-ring failures. These

observations are aligned with what we saw from the distribution and box plot charts in the Data section.

We took this analysis further by computing the odd ratio change of O-ring failure for a 10-unit increase in temperature or pressure. The odds of an O-ring failure is 0.374 times as likely, or about 63% less likely, for a 10 unit increase in temperature. For a 10-unit The pressure increase, the Odd Ratio is 1.089 times as likely as no pressure change, which indicates very little impact of pressure increase on the odds of O-ring failure.

As noted in the above Wald test, Pressure is not statistically significant. We further performed the Likelihood Ratio (LR) test using Anova function to evaluate the effect of each variable and determine whether Pressure should be kept or removed from the model.

```
Anova(mod1, test = "LR")
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: cbind(O.ring, Number - 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
```

In the the LR test, we stated a null hypothesis that $\beta_2 = 0$, or Pressure does not add significant explanatory power for the odds of O-ring failure, on top of Temperature in the model.

From this test, we obtained a p-value for a Chi-squared distribution of 0.0228 for Temperature and 0.2145 for pressure. Like the Wald test, the p-value for Pressure is above 0.05 (the 95% confidence level), so there is not enough evidence to reject the null hypothesis.

Considering both Wald and LR test results, we conclude that Pressure does not add much explanatory power to the model. Excluding it from the model is especially recommended in this case, given the small sample size of 23 observations. The drop in the degree of freedom is 5% when we add one additional explanatory variable. AIC score also decreases with one fewer variable and lower AIC is better. Keeping the model simple and easy to interpret would also help with the further study.

We removed Pressure from the second model version (mod2) below to simplify the model and noted that this removal does not have too much impact on temperature coefficient. Temperature coefficient is -0.1156 in mod2, instead of -0.098 in mod1.

3.2 Confidence Intervals (20 points)

In the analysis, we defined Temperature as our explanatory variable in the second model version (mod2). We performed the LR test using the anova function to determined whether a quadratic term is needed in the logistic regression model. See model summary and test results below.

```

# Estimate the logistic regression model with Temperature
mod2 <- glm(formula = cbind(O.ring, Number - O.ring) ~ Temp,
  family = binomial(link = "logit"),
  data = challenger)

# Estimate the logistic regression model including quadratic term
mod3 <- glm(formula = cbind(O.ring, Number - O.ring) ~ Temp + I(Temp^2),
  family = binomial(link = "logit"),
  data = challenger)

anova(mod2, mod3, test = "Chisq")

```

```

## Analysis of Deviance Table
##
## Model 1: cbind(O.ring, Number - O.ring) ~ Temp
## Model 2: cbind(O.ring, Number - 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 anova LR test yielded a p-value of 0.48, much higher than the 0.05 critical value, which means that there is little to no evidence that a quadratic term is statistically significant. Therefore, we will move forward with `mod2` without the quadratic term.

In the following analysis, we used our fitted `mod2` model to estimate the probability of O-ring failures and the number of failures for each launch conditional on the temperature at the launch. The plots below show the estimated results by temperature, with the 95% confidence interval.

From the above chart, odds of O-Rings failure declines as temperature increases, as β_1 is negative.

The 95% confidence intervals are wider for temperatures lower than 65F and narrower for temperatures greater than 65F. The estimated interval for temperature equal to 30F is about (4,6); this illustrates that we have high variability for the expected number of incidents at the low temperature. when temperature goes above 65F and increases towards 80F, our model estimated the number of O-ring failure declines close to zero.

The reason for wider confidence interval for lower than 65F is because we have fewer observations below 65F with the minimum temperature in the data set being 53F. The forecast error for the below-65F is larger.

Given the temperature of 31F for the Challenger launch in 1986, we used our model to estimate the probability of an O-ring failure and compute a corresponding confidence interval. As shown in the results below, the estimated probability of O-ring failure at 31F temperate is 0.82, with 95% confidence interval of 0.14 to 0.99. This is a very wide confidence interval, as we noted above related to the forecast error for below 65F.

To apply the inference procedures, we assume asymptotic properties of a large sample size, independence of each O-Ring failure and each failure has equal probability. This data set only has 23 observations which does not meet the large sample size assumption. As discussed in the prior section, we also raised concerns over the independence assumption.

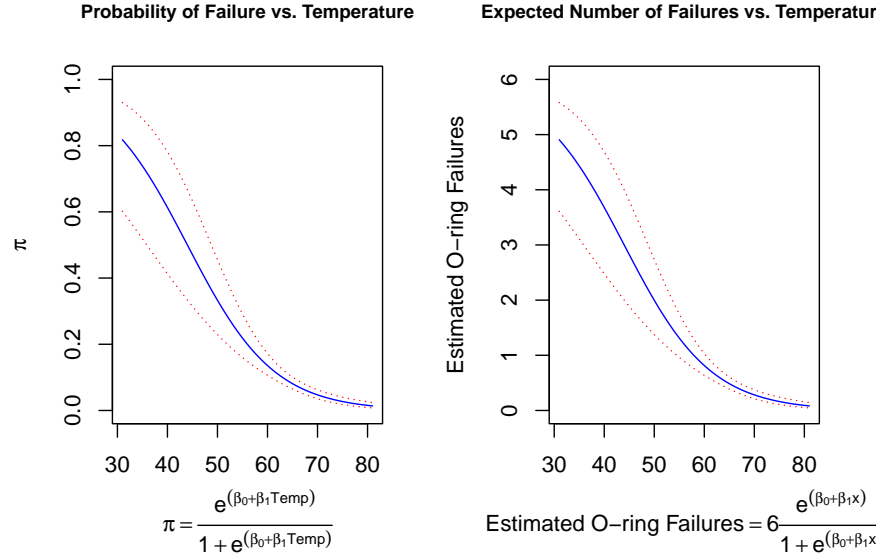


Figure 3: Temperature Effect on O-ring Failure Probability [Left] and Expected Failure Number [Right]

3.3 Bootstrap Confidence Intervals (30 points)

Statistical bootstrapping is a common method used in order to generate statistics with reasonable confidence. Since we are not able to perform the launching experiment for a large sample size, bootstrapping allows us to simulate such a process.

In this study, we performed 100-folds bootstrap of the challenger data set so that we can obtain 100 sets of O-ring failures at each temperature ranging from 10° to 100° Fahrenheit. It is worth mentioning that at each iteration, we re-sampled 23 observations from the original challenger data set with replacement. This sampling technique ensures that we do not have the same set of values in each iteration while allowing for repeating observations in the sample.

After the data set is re-sampled, we used this new data set to fit a model for the log odds of an O-ring failure using only temperature as explanatory variable. By fitting a new model for each re-sampled data set, we are able to obtain a slightly different coefficient that would ultimately yield a slightly different prediction at each iteration.

Finally, we record the probability of an O-ring failure as well as the 90% Wald confidence interval limits given an integer temperature between 10° to 100° Fahrenheit using the re-sampled model.

From the results obtained from the above procedures, we sliced the data on two temperatures values 31° F and 53° F to reference the temperature at which the Challenger space shuttle launch failed and the lowest launch temperature of the prior launches. From the two temperate data points, we generated the probability distribution histogram of our 100 bootstrap sampled as well as the Q-Q plot to illustrate the predicted probability deviations versus the theoretical Gaussian distribution.

Figure 4 below shows the probability distribution at 31° F. While the estimated O-ring failure probability distribution is sparse, it does show a high distribution concentration between 0.7 and 1.0. In contrast, the failure probability distribution at 53° F shown in Figure 5 is more Gaussian-like and aggregate at the low probability region between 0.0 and 0.4.

The Q-Q plots shown in Figure 4 and Figure 5 for both temperatures values further support this by showing the predicted failure probability versus the theoretical failure probability at each quantile. Note that the 53° F Q-Q plot shows an almost 1-1 match between the the quantiled probabilities while there is a clear deviation near the tail ends for the 31° F Q-Q plot.

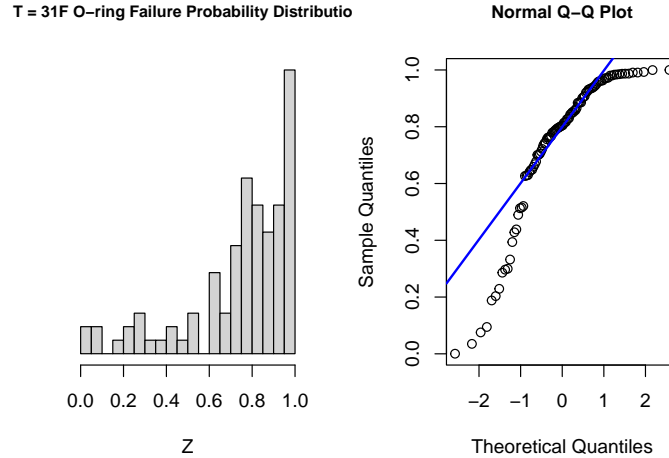


Figure 4: Bootstrap Distribution at T=31° F

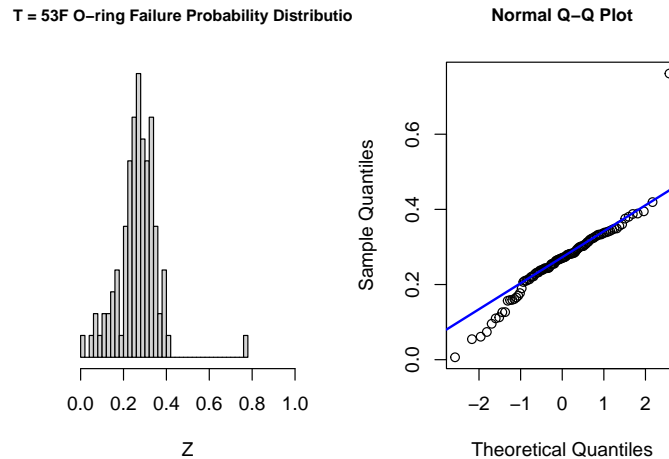


Figure 5: Bootstrap Distribution at T=53° F

It is reasonable to assume that if we're able to replicate the launch measurement at each temperature values 100 more times, we'll have a better idea of the probability of failure at each temperature. Using the bootstrapping method, we're able to better estimate the effect of launch temperature on the odds of an O-ring failure. In short, the temperature at which the Challenger space craft was launched at, 31° F has a really high probability of O-ring failure. Delaying the launch to a warmer temperature would have helped avoiding the failed outcomes.

3.4 Alternative Specification (10 points)

We considered the linear regression model as an alternative model selection. Since our statistical tests of the logistic regression models showed that pressure was not statistically significant, we only used temperature as an explanatory variable in the linear regression model to estimate the probability of failure. In this model (see the model form below), we defined the response variable as the number of O-rings failed divided by the total number of O-rings. We obtained a p-value of 0.013 for temperature, which is lowered than the stated 95% Wald confidence level, or alpha 0.05 and visualized by 1 asterisk. The negative coefficient for temperature suggests that at an increase in pressure decreases the probability of failure.

While the linear model generates reasonable coefficients, this approach is not suitable for this data set because the estimated probability of O-ring failure could fall outside of 0 and 1. Additionally, the linear regression model assumes that the probability of O-ring failure is linearly related to the explanatory variable, temperature in this case, that is not necessarily true. Figure 6 below also showed that it violates the homoskedasticity assumption, noting that the residuals versus fitted values line is not flat around 0. For those reasons, we will use a binary regression model for this data set, instead of the linear regression model.

```
model.linear <- lm(formula = I(O.ring / Number) ~ Temp, data = challenger)
```

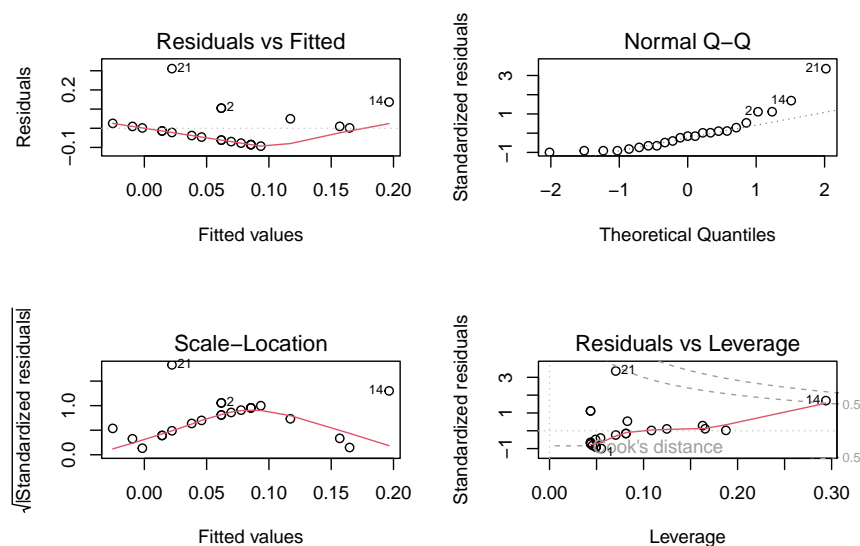


Figure 6: Linear Regression Residuals

4 Conclusions (10 points)

In conclusion, our final model (see Table 2 below) is the logistic regression model with O-ring success as the response variable and temperature as the explanatory variable. When temperature increases by 1-unit, the odds of O-ring failure will decrease by taking the exponent of the coefficient

of 0.116. When temperature increases by 10 degrees is, the odds of O-ring failure is 0.3147 times as large as the original temperature, which is about 69% decrease. Figure 3 above shows the estimated probability of O-ring failure and confidence interval at each temperature.

We conducted Wald and LR test to evaluate the importance of several explanatory variables (Temperature, Pressure, and quadratic term of Temperature). Our test results demonstrate that the logistic regression model with Temperature as the only explanatory variable is the most robust model. This answers the research question and concludes that temperature affects the odds of O-ring failure, but pressure does not have much effect. Given the negative coefficient of temperature, the probability of O-ring failure decreases as temperature increases. The confidence interval also dramatically narrows when temperature is greater than 65 degrees, compared to less than 65 degrees.

This observation is reinforced through parametric bootstrapping. With 100 iterations of bootstrapping and 23 observations per iteration, the results showed that at the temperature of 31F, the failure probability distribution is sparse but concentrates at a high probability region between 0.7 and 1.0. However, at 53F, the failure probability distribution is Gaussian-like and aggregates at a low probability region between 0.0 and 0.4. Therefore, delaying the launch to a warmer temperature could significantly decrease the probability of O-ring failure. Finally, we considered a linear regression model as an alternative model, but doing so would causes the probability of failure to be outside of the 0 to 1 range. Furthermore, the linear regression model would violates the assumption that the probability is linearly related to the explanatory variable as well as the homoskedasticity assumption.

Table 2: The Estimated Relationship Between O-Rings Failure and Temperature

Output Variable: Log Odds of O-Rings Failure	
Temperature	−0.116* p = 0.014
(Intercept)	5.085 p = 0.096
Observations	23
Log Likelihood	−15.823
Akaike Inf. Crit.	35.647
<i>Note:</i> *p<0.05; **p<0.01; ***p<0.001	