w271: Lab 1

Michael Berger, John Gao, and Thomas Hamnett

Lab 1: Investigation of the 1989 Space Shuttle Challenger Accident

1. Introduction

In this paper, we will examine the probability of an O-ring failure in space shuttle launches. We will use data collected by Dalal, Fowlkes and Hoadley from their 1989 paper "Risk Analysis of the Space Shuttle: Pre-Challenger Prediction of Failure", in order to model probability of failure as a function of temperature at launch and combustion pressure.

The key question being asked is, if O-ring failure can be reliably modeled using launch temperature and/or combustion pressure. The question is motivated by the Challenger shuttle tragedy, in which initially it was thought O-ring failure resulted in its tragic crash, and the authors sought to conduct an analysis that could help prevent future crashes.

In order to inform this analysis, we conducted a thorough EDA of the shuttle data, which helped inform our model specification. In the end, we found that a binary logistic regression modeling any O-ring failure vs no O-ring failure, as a function of launch temperature only (and not combustion pressure), was the most appropriate model given the data. Reasons for a binary model against a binomial model include stricter independence assumptions within observations for the binomial model, which are not met, and limited data on multiple O-ring failures (so using a binary model does not lose much information). Reasons for using temperature only include analysis of variance between model specifications with and without a pressure term that show addition of pressure term does not show significant difference from the temperature only model.

Using our model, we found the estimated probability of a failure of at least one O-ring at the average temperature at launch (~70F) is 24.83%, while at 31 degrees it is 99.96%. This indicates that at low temperatures approaching freezing, it is inadvisable to move forward with a shuttle launch. The analysis can provide guidance in decision making on whether to proceed with a launch given launch temperature.

2. Data Loading, Analysis and Exploratory Data Analysis (EDA)

Loading Libraries and Data and Analysis of Data

First we load the libraries, then the data. We inspect the data and conduct EDA.

```
# Libraries
library(knitr)
library(stargazer)
library(car)
library(dplyr)
library(Hmisc)
library(ggplot2)
library(gridExtra)
# We use the gridExtra library to call the grid.arrange() function
# Details: https://cran.r-project.org/web/packages/gridExtra/gridExtra.pdf
```

```r
library(kableExtra)
# We use the kableExtra library to format tables
# Details: https://cran.r-project.org/web/packages/kableExtra/kableExtra.pdf
opts_chunk$set(tidy.opts=list(width.cutoff=60),tidy=TRUE)

# Loading data
challenger_raw <- read.csv('challenger.csv', sep = ',', header = TRUE)
# Inspecting data
str(challenger_raw)
```

```
## 'data.frame':    23 obs. of  5 variables:
##  $ Flight  : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Temp    : int  66 70 69 68 67 72 73 70 57 63 ...
##  $ Pressure: int  50 50 50 50 50 50 100 100 200 200 ...
##  $ O.ring  : int  0 1 0 0 0 0 0 0 1 1 ...
##  $ Number  : int  6 6 6 6 6 6 6 6 6 6 ...
```

```r
kable(summary(challenger_raw), "latex", booktabs = T) %>%
  kable_styling(font_size = 8)
```

Flight	Temp	Pressure	O.ring	Number
Min. : 1.0	Min. :53.00	Min. : 50.0	Min. :0.0000	Min. :6
1st Qu.: 6.5	1st Qu.:67.00	1st Qu.: 75.0	1st Qu.:0.0000	1st Qu.:6
Median :12.0	Median :70.00	Median :200.0	Median :0.0000	Median :6
Mean :12.0	Mean :69.57	Mean :152.2	Mean :0.3913	Mean :6
3rd Qu.:17.5	3rd Qu.:75.00	3rd Qu.:200.0	3rd Qu.:1.0000	3rd Qu.:6
Max. :23.0	Max. :81.00	Max. :200.0	Max. :2.0000	Max. :6

```r
sum(complete.cases(challenger_raw))
```

```
## [1] 23
```

```r
head(challenger_raw, 4)
```

```
##   Flight Temp Pressure O.ring Number
## 1      1   66       50      0      6
## 2      2   70       50      1      6
## 3      3   69       50      0      6
## 4      4   68       50      0      6
```

Based on the `complete.cases()` function, we see that no variable in the data set has missing values. There are 23 observations, representing space shuttle starts, with five variables *Flight*, *Temp*, *Pressure*, *O.ring*, and *Number*. We will not use *Flight* as it appears to serve as an index of each flight record and we will not use *Number* as it represents the number of O-rings, which is constant at 6 for each flight.

We see that there are two potential explanatory variables included in this data: - *Temp*: Representing the temperature at the start of the space shuttle in degrees Fahrenheit. - *Pressure*: We note from the paper that *Pressure* is defined as the pressure at which the O-rings of a shuttle mission were tested.

Both are potential determinants of O.ring. For the time being, we completely assume away the issue of omitted variable bias.

The response (or dependent) variable of interest, number of O-rings that failed, is denoted as *O.ring*. *O.ring* takes on the values 0, 1, and 2. Hence, we observe maximum 2 failures out of 6 possible failures per observation.

As the explanatory variable represents *failure* or *non-failure* with different levels of failure (1 - 6), we can model this using a binomial logistic regression model estimating the number of O-ring failures. We can also create a binary *Incident* variable representing *failure* and *non-failure* and then fit a binary logistic regression model to estimate probability of failure.

We further note from `str()` that only three different pressure values are present, namely 50, 100 and 200.

We will next remove *Flight* and *Number*, create a categorical variable *Fpressure* for the three levels of *Pressure* and create a binary incident variable *Incident*. We also create a *Foring* variable to reflect the number of O-ring failures as factors for graphing purposes.

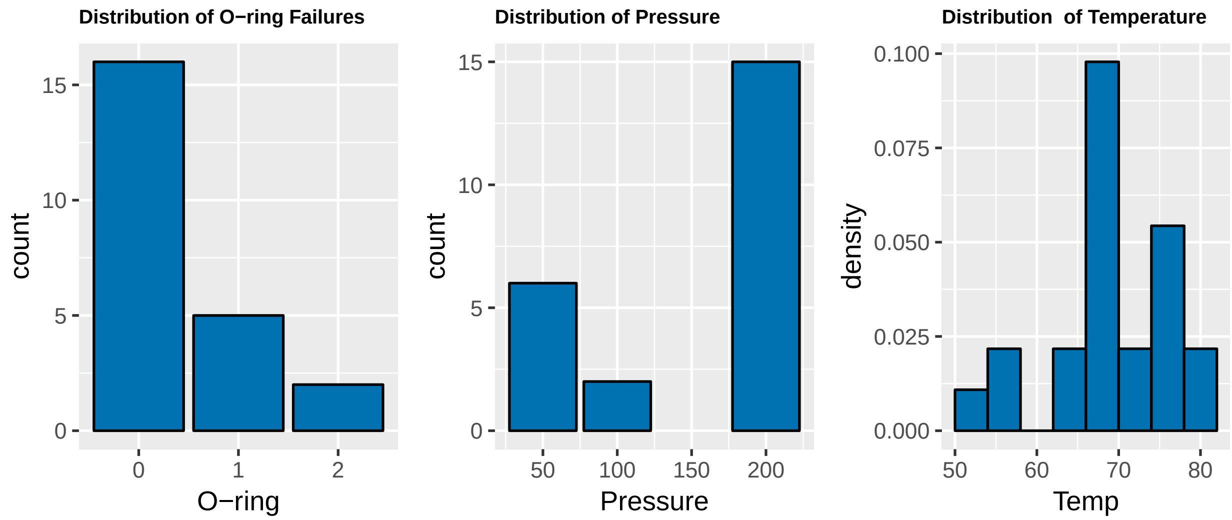
```
# Making a copy of the raw data to perform the
# transformations on
challenger <- challenger_raw
challenger$Incident <- factor(challenger$O.ring > 0)
challenger$Foring <- factor(challenger$O.ring)
challenger$Fpressure <- factor(challenger$Pressure)
challenger$Flight <- NULL
challenger$Number <- NULL
```

Exploratory Data Analysis

Next, we perform an exploratory data analysis.

```
# Distribution of O-ring failure
plot1 <- ggplot(challenger, aes(x = Foring)) + geom_bar(fill = "#0072B2",
    colour = "black") + ggtitle("Distribution of O-ring Failures") +
    xlab("O-ring") + theme(plot.title = element_text(lineheight = 1,
    face = "bold", size = (8)))
# Distribution of temperature
plot2 <- ggplot(challenger, aes(x = Temp)) + geom_histogram(aes(y = ..density..),
    fill = "#0072B2", colour = "black", bins = 8) + ggtitle("Distribution  of Temperature") +
    theme(plot.title = element_text(lineheight = 1, face = "bold",
        size = (8)))
# Distribution of pressure
plot3 <- ggplot(challenger, aes(x = Pressure)) + geom_bar(fill = "#0072B2",
    colour = "black") + ggtitle("Distribution of Pressure") +
    theme(plot.title = element_text(lineheight = 1, face = "bold",
        size = (8)))

grid.arrange(plot1, plot3, plot2, nrow = 1)
```

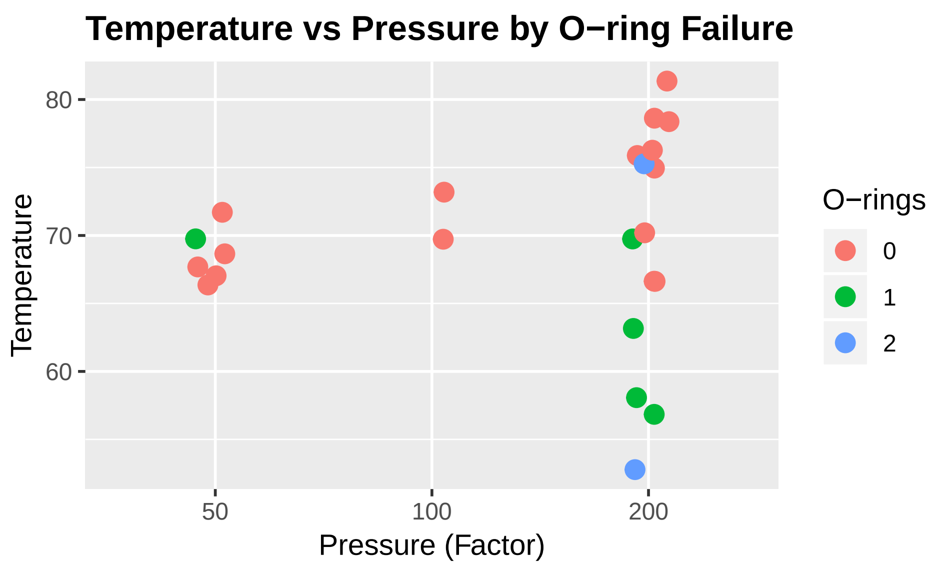


Univariate analysis

- *O.ring*: *O.ring* takes on a value in our sample of 0, 1, or 2 to indicate the number of O-ring failures in a flight (presumably this value could be as high as 6 if all 6 O-rings failed). 16 of 23, or 70% have O.ring = 0, 5 have O.ring = 1 (22%) and 2 have O.ring = 2 (9%).
- *Temp*: *Temp* indicates temperature in Fahrenheit at time of launch. In the data, minimum is 53 and maximum is 81. Mean is 69.6 while median is 70, indicating roughly half are above and below the mean. Temp has standard deviation of 7.1, and the middle 50% of values are between 67 and 75. Skew appears negative, potentially indicating a non-normal distribution.
- *Pressure*: *Pressure* indicates the psi of the O-ring test conducted prior to flights. Pressure takes on values of 50, 100, and 200. 6 of 23, or 60% have Pressure = 0, 2 have Pressure = 100 (26%) and 15 have Pressure = 200 (65%).

Multivariate Analysis

```
# Temperature vs Pressure by O-ring failure Pressure plotted
# as a factor for visual clarity
ggplot(challenger, aes(x = Fpressure, y = Temp, color = Foring)) +
    geom_jitter(size = 3, width = 0.1) + labs(color = "O-rings") +
    ggtitle("Temperature vs Pressure by O-ring Failure") + theme(plot.title = element_text(lin
    face = "bold")) + xlab("Pressure (Factor)") + ylab("Temperature")
```



- *Incident vs Temp*: Above plot shows that below the temperature of 65, all flights had at least one O-ring failure. Two other O-ring failures occurred at a temperature of 70 and one at 75 (with two failures). This shows while *Temp* does not completely separate the *Incident* variable, below 65F each flight in the sample had at least one O-ring failure.
- *Temp vs Pressure*: Looking at *Temp* vs *Pressure*, with points colored by O-ring failure, surfaces several insights. At 50 and 100 psi, the temperature range was between 65-75, with only one incident. At 200 psi, the temperature range is from 53 to 81 (both min an max of range), and there was 1 incident at 75 temperature and 1 at 70 temperature. The temperature range below 65 at 200 psi all had incidents (4 incidents).

Tabular Analysis

```
# Observations by O-ring Failure, Counts and %
toring <- table(challenger$O.ring)
ptoring <- round(prop.table(table(challenger$O.ring)), 2)
kable(cbind(toring, ptoring), "latex", booktabs = T) %>% kable_styling(font_size = 8)
```

	toring	ptoring
0	16	0.70
1	5	0.22
2	2	0.09

```
# Observations by Incident Occurrance, Counts and %
tinc <- table(challenger$Incident)
ptinc <- round(prop.table(table(challenger$Incident)), 2)
kable(cbind(tinc, ptinc), "latex", booktabs = T) %>% kable_styling(font_size = 8)
```

	tinc	ptinc
FALSE	16	0.7
TRUE	7	0.3

```r
# Observations by Pressure Level, Counts and %
tpres <- table(challenger$Pressure)
ptpres <- round(prop.table(table(challenger$Pressure)), 2)
kable(cbind(tpres, ptpres), "latex", booktabs = T) %>% kable_styling(font_size = 8)
```

	tpres	ptpres
50	6	0.26
100	2	0.09
200	15	0.65

```r
# Incident by Pressure Range, Counts and % Total
xincpr <- xtabs(~Incident + Pressure, data = challenger)
pxincpr <- round(prop.table(xtabs(~Incident + Pressure, data = challenger)),
    2)
kable(cbind(xincpr, pxincpr), "latex", booktabs = T) %>% kable_styling(font_size = 8)
```

	50	100	200	50	100	200
FALSE	5	2	9	0.22	0.09	0.39
TRUE	1	0	6	0.04	0.00	0.26

```r
# Incident % Within Pressure Level
kable(round(prop.table(xtabs(~Incident + Pressure, data = challenger),
    2), 2), "latex", booktabs = T) %>% kable_styling(font_size = 8)
```

	50	100	200
FALSE	0.83	1	0.6
TRUE	0.17	0	0.4

```r
# Create factor variable to examine Temp vs Incident
challenger$ftemp <- cut2(challenger$Temp, cuts = c(65, 74), minmax = T)
# Incident by Temp Range, Counts and % Total
xinctemp <- xtabs(~Incident + ftemp, data = challenger)
pxinctemp <- round(prop.table(xtabs(~Incident + ftemp, data = challenger)),
    2)
kable(cbind(xinctemp, pxinctemp), "latex", booktabs = T) %>%
    kable_styling(font_size = 8)
```

	[53,65)	[65,74)	[74,81]	[53,65)	[65,74)	[74,81]
FALSE	0	10	6	0.00	0.43	0.26
TRUE	4	2	1	0.17	0.09	0.04

```r
# Incident % Within Temp Range
kable(round(prop.table(xtabs(~Incident + ftemp, data = challenger),
    2), 2), "latex", booktabs = T) %>% kable_styling(font_size = 8)
```

	[53,65)	[65,74)	[74,81]
FALSE	0	0.83	0.86
TRUE	1	0.17	0.14

The sample proportion of 0 is 70% (or 16 flights in the sample), of 1 is 22% (5 flights), and of 2 is 9% (2 flights). In total, *Incident* is *TRUE* 30% of the time (7 flights).

A tabular view against three temp ranges, 53-65, 65-74, and 74-81, show 4 incidents at 53-65 (100% of range), 2 incidents at 65-74 (17% of range), and 1 incident at 75-81 (14% of range).

Finally, as noted earlier, *Pressure* represents the psi of the test conducted pre-flight. The paper states that the pressure was increased on later flights since lower pressures might not penetrate the putty. We created an ordinal factor variable for pressure to represent the three levels of pressure. At 50 psi, there were 6 flights, 1 of which had an incident (17% among 50 psi). At 100 psi, there were 2 flights, both of which without incident. At 200 psi, there were 15 flights, 6 of which had an incident (40%). This shows that while *Pressure* does not completely separate the *Incident* variable, at 200 psi the rate of failure was higher than for the lower pressures.

We now continue to the model building phase and the questions to be answered.

2. Answer question 4 and 5 on Chapter 2 (page 129 and 130) of Bilder and Loughin's *"Analysis of Categorical Data with R"*

Question 4:

a. Discussion of why the assumption of independence of each O-ring is necessary and the potential problems with this assumption

The binomial logistic model fit on the data by Dalal et al., treating each O-ring as a single observation, needs to assume that the observations are independent and that the error terms are independent and identically distributed. This is why the failure of a single O-ring needs to be assumed to be independent. The failure of one O-ring should not provide any information about the failure of any other O-ring.

The issue with this assumption is that at each start of a shuttle, six O-rings are exposed to a similar heat and pressure. As heat and pressure are assumed to be causes for failure, each O-ring is exposed to a similar condition and hence the failure of the same O-rings at a single space shuttle start is not independent from each other. This is the reason why we decided against a binomial logistic regression model and in favor of a binary logistic regression model.

It is worth noting that even with a binary logistic regression model, there is an independence assumption between observations. However, it stands to reason that with each subsequent launch, there would have been adjustments made based on any problems observed from prior launches. So even in the less restrictive binary logistic regression model, there may be potential violations of the independence assumption.

b. Estimating the logistic regression model using the explanatory variables in a linear form

Based on the issue of independence weighted against the comparable minor loss of information, we came to the conclusion that a binary logistic regression model on whether at least one O-ring failed would be more suitable than a binomial logistic regression on the count data of O-ring failures. Hence, we defined a variable $Incident = I(Y > 0)$, with Y denoting the random variable representing O-ring failures at a space shuttle start and $I()$ being an indicator function. *Incident* captures then the probability of at least one O-ring failure vs no O-ring failure. As noted by Dalal et al. (1989), p. 949, when we want to derive the probability of one single O-ring failure, denoted as $P(Y = 1)$, and if we accept the assumption of independence between the single O-ring failures, then we can compute $P(Y > 0) = 1 - (1 - P(Y = 1))^6$ or equivalently $P(Y = 1) = 1 - (1 - P(Y > 0))^{1/6}$. We keep this transformation in mind when looking at the expected number of O-ring failures.

We now estimate the binary logistic regression model in the form:

$$logit(Incident) = \beta_0 + \beta_1 Temp + \beta_2 Pressure + \epsilon$$

```r
# binary model
challenger_binary.glm <- glm(Incident ~ Temp + Pressure, family = binomial(link = logit),
    data = challenger)
```

Our estimated binary logistic regression model is:

$$logit(\hat{Incident}) = 13.292 - 0.229 \times Temp + 0.010 \times Pressure$$

As comparison, we also fit the binomial logistic regression model as used in the Dalal et al. paper.

```r
# binomial model
challenger_binomial.glm <- glm(cbind(O.ring, 6 - O.ring) ~ Temp +
    Pressure, family = binomial(link = logit), data = challenger)
```

We summarize the estimated parameters and the associated statistics in below regression output table. Here, we report the Wald test statistics as in the `summary()` function.

```r
stargazer(challenger_binary.glm, challenger_binomial.glm, type = "latex",
    column.labels = c("Binary Model", "Binomial Model"), title = "Logistic Regression Models P
    font.size = "footnotesize", dep.var.labels = c("O-Ring Failure Incident",
        "Number of O-ring Failures"), omit.stat = "f", intercept.bottom = FALSE,
    align = TRUE)
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu % Date and time: Mon, Feb 04, 2019 - 5:38:16 PM % Requires LaTeX packages: dcolumn

We note that the intercept in the binary logistic regression model is much higher than in the binomial model and also slightly significant at the 10% level. Both models also show that *Pressure* is not significant, while *Temp* is significant at the 5% level.

In the following we will only use the estimated binary logistic regression model due to the issue of the independence assumption in the binomial logistic regression model and the comparatively only minor loss of information.

c. Performing LRTs to judge importance of variables in the model

We use the `Anova()` function to perform the Likelihood Ratio Tests (LRT).

Table 1: Logistic Regression Models Predicting O-ring Failure

	Dependent variable:	
	O-Ring Failure Incident Binary Model	Number of O-ring Failures Binomial Model
	(1)	(2)
Constant	13.292*	2.520
	(7.664)	(3.487)
Temp	−0.229**	−0.098**
	(0.110)	(0.045)
Pressure	0.010	0.008
	(0.009)	(0.008)
Observations	23	23
Log Likelihood	-9.391	-15.053
Akaike Inf. Crit.	24.782	36.106
Note:		*p<0.1; **p<0.05; ***p<0.01

```r
Anova(challenger_binary.glm, test = "LR")
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: Incident
##          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
```

We note that *Temp* is significant to the 1% level for the binary logistic regression model, while *Pressure* is not significant even at the 10% level. Compared to the Wald test above, we note that the LRT shows a higher significance for temperature.

d. Reason for removing the `Pressure` variable and potential problems

The authors removed the *Pressure* variable as it did not show significance (see Dalal et al.(1989), p. 948).

The problem with removing this variable **AFTER RUNNING THE TEST** is that under the frequentist approach the hypothesis was made and then the test was run on a data sample. After evaluating the test and evaluating the hypothesis, the same sample should not be used to establish a new hypothesis and specify a different model without *Pressure*. However, it appears that the authors have done this. This proves problematic and influences the validity of the test statistics.

Question 5:

a. Estimating one-variable-logit model with temperature only

We now estimate the binary model with *Temp* as only explanatory variable. Our model specification is as follows:

$$logit(Incident) = \beta_0 + \beta_1 Temp + \epsilon$$

```r
# binary model - one variable
challenger_binary_1v.glm <- glm(Incident ~ Temp, family = binomial(link = logit),
    data = challenger)
```

```r
stargazer(challenger_binary_1v.glm, challenger_binary.glm, column.labels = c("Binary Model with
    "Binary Model with 2 Variables"), title = "Logistic Regression Models Predicting O-ring Fa
          Temp Only vs Temp and Pressure",
    font.size = "footnotesize", dep.var.labels = c("O-Ring Failure Incident"),
    type = "latex", omit.stat = "f", intercept.bottom = FALSE,
    align = TRUE)
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu % Date and time: Mon, Feb 04, 2019 - 5:38:16 PM % Requires LaTeX packages: dcolumn

Table 2: Logistic Regression Models Predicting O-ring Failure: Temp Only vs Temp and Pressure

	Dependent variable:	
	O-Ring Failure Incident	
	Binary Model with 1 Variable	Binary Model with 2 Variables
	(1)	(2)
Constant	15.043^{**}	13.292^{*}
	(7.379)	(7.664)
Temp	-0.232^{**}	-0.229^{**}
	(0.108)	(0.110)
Pressure		0.010
		(0.009)
Observations	23	23
Log Likelihood	-10.158	-9.391
Akaike Inf. Crit.	24.315	24.782
Note:		*p<0.1; **p<0.05; ***p<0.01

Our fitted one-variable binary logistic regression model is:

$$logit(\hat{Incident}) = 15.043 - 0.232 \times Temp$$

When comparing the one variable to the two variable binary logistic regression model, we note that the intercept became more significant. Otherwise, the estimated parameters stayed nearly the same with only slight changes.

b. Constructing two Plots: π vs `Temp` and Expected number of failures vs `Temp`

We use the binary logistic regression model. However, the authors of the paper used the binomial model instead. Due to the reasons as outlined above, we think the binary model is the better reference model.

As outlined under 4a. we use the transformation $\hat{P}(Y = 1) = 1 - (1 - \hat{P}(Y > 0))^{1/6} = 1 - (1 - \frac{e^{15.043 - 0.232 \times Temp}}{1 + e^{15.043 - 0.232 \times Temp}})^{1/6}$. To use this transformation, however, we need to assume independence between single O-ring failures. Then we can compute the expected number of failures by using the expected value formula for a binomial distribution, $E(Y) = n\hat{P}(Y = 1)$, with $n = 6$ representing the number of O-rings per flight.

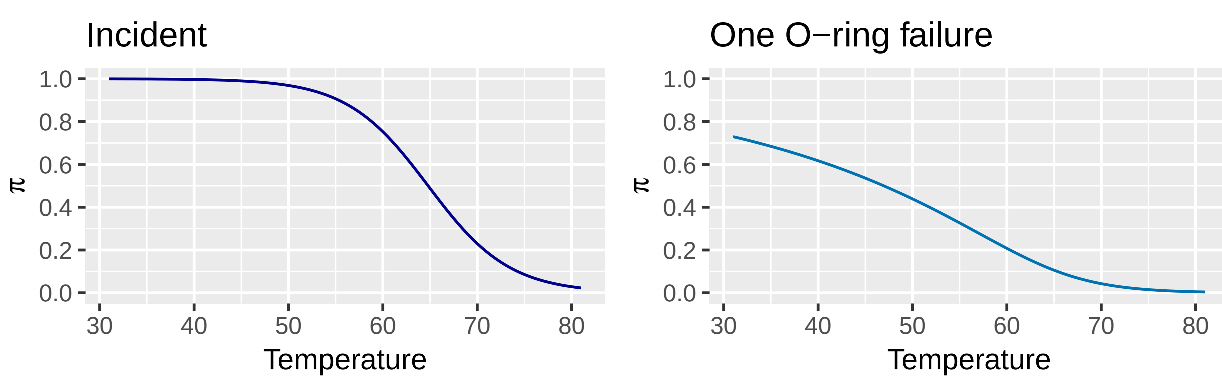
Below we plot two graphs, one with the probability of an incident at different temperature levels (at least one O-ring failure) and the probability of one O-ring failure at different temperatures.

```r
alpha <- challenger_binary_1v.glm$coef[1]
beta <- challenger_binary_1v.glm$coef[2]

plot1 <- ggplot(data.frame(x = c(31, 81), y = c(0, 1)), aes(x,
    y)) + stat_function(fun = function(x) exp(alpha + beta *
    x)/(1 + exp(alpha + beta * x)), geom = "line", color = "darkblue") +
    ggtitle("Incident") + xlab("Temperature") + ylab(expression(pi)) +
    scale_y_continuous(breaks = c(0, 0.2, 0.4, 0.6, 0.8, 1),
        limits = c(0, 1))

plot2 <- ggplot(data.frame(x = c(31, 81), y = c(0, 1)), aes(x,
    y)) + stat_function(fun = function(x) 1 - (1 - exp(alpha +
    beta * x)/(1 + exp(alpha + beta * x)))^(1/6), geom = "line",
    color = "#0072B2") + ggtitle("One O-ring failure") + xlab("Temperature") +
    ylab(expression(pi)) + scale_y_continuous(breaks = c(0, 0.2,
    0.4, 0.6, 0.8, 1), limits = c(0, 1))

grid.arrange(plot1, plot2, nrow = 1)
```

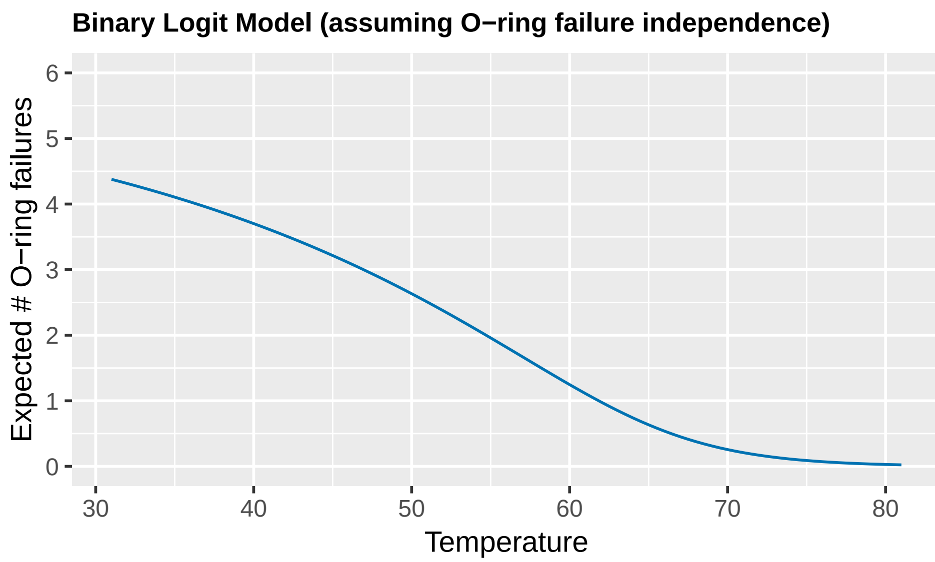


We note from the graphs that the probability of an incident remains very high for temperatures below 65F.

To plot the expected number of failures by using the binary logistic regression model, we use the relationship and transformation noted above. Please note that we need to assume O-ring failure independence.

```
alpha <- challenger_binary_1v.glm$coef[1]
beta <- challenger_binary_1v.glm$coef[2]
n <- 6

ggplot(data.frame(x = c(31, 81), y = c(0, 1)), aes(x, y)) + stat_function(fun = function(x) (1
    (1 - exp(alpha + beta * x)/(1 + exp(alpha + beta * x)))^(1/6)) *
    n, geom = "line", color = "#0072B2") + ggtitle("Binary Logit Model (assuming O-ring failur
    xlab("Temperature") + ylab("Expected # O-ring failures") +
    theme(plot.title = element_text(lineheight = 1, face = "bold",
        size = (10))) + scale_y_continuous(breaks = c(0, 1, 2,
    3, 4, 5, 6), limits = c(0, 6))
```



We note that the graph suggests that the expected number of O-ring failures exceeds 1 after the temperature falls below 60F. At 31F we would expect more than 4 O-ring failures. This seems to indicate a high risk of failure in case the temperature drops below this threshold.

c. Computing 95% Wald Interval, adding it to the plot and discussing why the interval is wider for lower temperatures

Next we compute the 95% Wald confidence intervals for an incident as well as for a single O-ring failure and add the computed intervals to the plots.

```
# Creating a function
CI.func <- function(new_data, model_object, alpha, xpo) {
    # Obtain the linear predictor
    linear.pred <- predict(object = model_object, newdata = new_data,
        type = "link", se = TRUE)
    # Computing pi.hat
    pi.hat <- 1 - (1 - exp(linear.pred$fit)/(1 + exp(linear.pred$fit)))^(1/xpo)
```

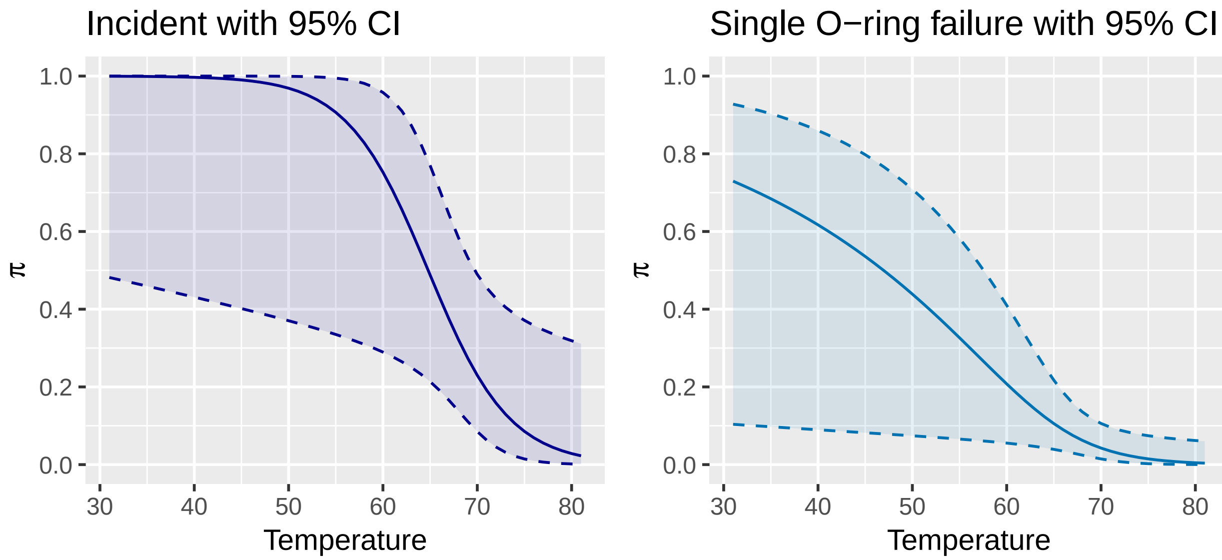
```r
    # Computing Wald Confidence Interval
    CI.lin.lower <- linear.pred$fit - qnorm(p = 1 - alpha/2) *
        linear.pred$se
    CI.lin.upper <- linear.pred$fit - qnorm(p = alpha/2) * linear.pred$se
    CI.pi.lower <- 1 - (1 - exp(CI.lin.lower)/(1 + exp(CI.lin.lower)))^(1/xpo)
    CI.pi.upper <- 1 - (1 - exp(CI.lin.upper)/(1 + exp(CI.lin.upper)))^(1/xpo)
    # Returning result as list
    list(pi_est = pi.hat, lower = CI.pi.lower, upper = CI.pi.upper)
}

# Plotting the confidence level bands, with shading
alpha <- challenger_binary_1v.glm$coef[1]
beta <- challenger_binary_1v.glm$coef[2]

ci.data.func <- function(xpo2) {
    ci.data.case = data.frame(x = c(31:81))
    ci.data.case = data.frame(x = c(31:81), mean = sapply(ci.data.case$x,
        FUN = function(x) {
            1 - (1 - exp(alpha + beta * x)/(1 + exp(alpha + beta *
                x)))^(1/xpo2)
        }), lower = sapply(ci.data.case$x, FUN = function(x) {
        CI.func(new_data = data.frame(Temp = x), model_object = challenger_binary_1v.glm,
            alpha = 0.05, xpo = xpo2)$lower
    }), upper = sapply(ci.data.case$x, FUN = function(x) {
        CI.func(new_data = data.frame(Temp = x), model_object = challenger_binary_1v.glm,
            alpha = 0.05, xpo = xpo2)$upper
    }))
    return(ci.data.case)
}

ciplot <- function(ci.data, ci.title, color) {
    ggplot(ci.data, aes(x = x, y = mean)) + geom_line(aes(y = mean),
        color = color) + geom_line(aes(y = lower), color = color,
        linetype = "dashed") + geom_line(aes(y = upper), color = color,
        linetype = "dashed") + geom_ribbon(aes(ymin = lower,
        ymax = upper), fill = color, alpha = 0.1) + ggtitle(ci.title) +
        xlab("Temperature") + ylab(expression(pi)) + scale_y_continuous(breaks = c(0,
        0.2, 0.4, 0.6, 0.8, 1), limits = c(0, 1))
}

grid.arrange(ciplot(ci.data.func(xpo2 = 1), "Incident with 95% CI",
    "darkblue"), ciplot(ci.data.func(xpo2 = 6), "Single O-ring failure with 95% CI",
    "#0072B2"), nrow = 1)
```



We see that the confidence intervals are much higher with lower temperatures. This is likely due to having very few observations for lower temperatures. Despite our only four observations at temperatures below 65F showing at least 1 O-ring failure, the limited number of data points leads to larger CI for lower temperatures. The same holds true for the probability of a single O-ring failure.

At higher temperatures, two factors lead to lower CI ranges. One is the overall larger number of data points at higher temperatures, and the second is that at higher temperatures, most of the data points show no incident. Therefore, there is sufficient evidence for a smaller CI at higher temperatures than for lower temperatures.

d. Computing confidence interval for probability of failure at 31 degrees and discussion of assumptions need to be made

We use the previously defined function to compute the confidence interval for the probability of an incident at 31F, the temperature at the start of the Challenger in 1986.

```r
# Using function defined
round(data.frame(temperature = 31, CI.func(new_data = data.frame(Temp = 31),
    model_object = challenger_binary_1v.glm, alpha = 0.05, xpo = 1)),
    4)
```

```
##   temperature pi_est  lower upper
## 1          31 0.9996 0.4816     1
```

The estimated probability of failure of at least one O-ring at a temperature of 31 degree is 99.96%, with the 95% confidence interval being $[0.4816, 1]$. This indicates that the risk of an incident is very high.

The assumptions to be made in order for the inference procedure to apply are the following:

- Each space shuttle mission (our n trials) are identical trials: This seems not entirely plausible, as O-rings changed during the mission (higher pressure for testing).
- Each space mission has two possible outcomes (success or failure): This is not the case here, as more than 1 O-ring could fail. However, as we have seen during the EDA, the maximum O-ring

failures were 2. Hence, we think that the lost information by applying a dummy incident variable is low.

- The trials are independent of each other: This seems not plausible, as the experience from one start (e.g. technical defects) informed and influenced the next mission.
- The probability of an O-ring incident π is constant for each mission: This again seems not plausible, as different pressure tested O-rings were used.
- The random variable $w = O.ring$ represents the number of failures: This is true.

Hence, we think most assumptions are at least not entirely met, which might influence the validity of the statistical test.

e. Using parametric bootstrap to compute confidence interval

Next, we implement the suggested parametric bootstrap procedure to estimate a 90% confidence interval.

```
set.seed(2019)

bootstrap <- function(n, temp, model_object, simulation) {
    result = list()
    for (i in 1:simulation) {
        new_data <- data.frame(Temp = sample(challenger$Temp,
            size = n, replace = TRUE))
        linear.pred <- predict(object = model_object, newdata = new_data,
            type = "link")
        new_data$Prob <- exp(linear.pred)/(1 + exp(linear.pred))
        new_data$Incident <- round(rbinom(n = n, size = 1, prob = new_data$Prob),
            0)

        new_model <- glm(Incident ~ Temp, family = binomial(link = logit),
            data = new_data)
        new.pred <- predict(object = new_model, newdata = data.frame(Temp = temp),
            type = "link")
        prob <- exp(new.pred)/(1 + exp(new.pred))
        result[[i]] <- prob
    }
    quantile(unlist(result, use.names = FALSE), probs = c(0.05,
        0.95), na.rm = TRUE)
}

print("Temperature = 31F:")
```

```
## [1] "Temperature = 31F:"
```

```
suppressWarnings(bootstrap(n = 23, temp = 31, model_object = challenger_binary_1v.glm,
    simulation = 10000))
```

```
##        5%       95%
## 0.9472235 1.0000000
```

```r
print("Temperature = 71F:")
```

```
## [1] "Temperature = 71F:"
```

```r
suppressWarnings(bootstrap(n = 23, temp = 71, model_object = challenger_binary_1v.glm,
    simulation = 10000))
```

```
##         5%        95%
## 0.01039805 0.35610246
```

Given the above specification with `replace=TRUE`, we need to note that complete separation issue might most likely occur, resulting in `NA`s for the respective simulation run. Hence we suppressed the warnings and calculated the means by using `na.rm`.

The calculated 90% confidence intervals for a temperature of 31F indicate a probability of at least one O-ring failure to be between $[0.9472, 1]$, compared to a confidence interval of $[0.0103, 0.3561]$ for a temperature of 71F. For 31F we see that the confidence interval is much tighter compared to the Wald CI.

f. Determine if quadratic term is needed

Next, we fit a model with a quadratic temperature term and use the `anova()` function to perform a LRT regarding whether the quadratic term is significant. Our model with quadratic term is specified as follows:

$$logit(Incident) = \beta_o + \beta_1 Temp + \beta_3 Temp^2 + \epsilon$$

```r
challenger_binary_1v_quadratic.glm <- glm(Incident ~ Temp + I(Temp^2),
    family = binomial(link = logit), data = challenger)
summary(challenger_binary_1v_quadratic.glm)
```

```
##
## Call:
## glm(formula = Incident ~ Temp + I(Temp^2), family = binomial(link = logit),
##     data = challenger)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9614  -0.6608  -0.4735   0.1781   2.1185
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 78.48317   76.67353   1.024    0.306
## Temp        -2.09430    2.18479  -0.959    0.338
## I(Temp^2)    0.01359    0.01553   0.875    0.381
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 28.267  on 22  degrees of freedom
```

```
## Residual deviance: 19.389  on 20  degrees of freedom
## AIC: 25.389
##
## Number of Fisher Scoring iterations: 6
```

```
anova(challenger_binary_1v.glm, challenger_binary_1v_quadratic.glm,
    test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: Incident ~ Temp
## Model 2: Incident ~ Temp + I(Temp^2)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        21     20.315
## 2        20     19.389  1  0.92649   0.3358
```

The estimated binary logistic regression model with quadratic term is:

$$logit(\hat{Incident}) = 78.48 - 2.09 \times Temp + 0.01 \times Temp^2$$

We note that the LRT yields that the quadratic term is not significant. Hence, it is not needed in the model.

3. In addition to the questions in Question 4 and 5, answer the following questions:

a. Interpret the main result of your final model in terms of both odds and probability of failure

We set up the code to compute the odds of failure by a change in temperature in degree Fahrenheit. We use $c = -10$ as unit of change, hence a reduction of temperature by 10F.

```
# odds
c <- -10  # decrease in temperature by 10 degrees sounds reasonable
exp(c * coef(challenger_binary_1v.glm)[2])
```

```
##     Temp
## 10.19225
```

Decreasing the temperature by 10 degrees increases the odds of at least one O-ring failure by 10.19 times.

Next, we compute the probabilities of at least one O-ring failure. We compute the probability of failure at the mean temperature level of the space starts and once at the 31F level.

```
temp_1 <- mean(challenger$Temp)
temp_2 <- 31

# probability of failure
exp(challenger_binary_1v.glm$coef[1] + temp_1 * challenger_binary_1v.glm$coef[2])/(1 +
    exp(challenger_binary_1v.glm$coef[1] + temp_1 * challenger_binary_1v.glm$coef[2]))
```

```
## (Intercept)
##   0.2483279
```

```
exp(challenger_binary_1v.glm$coef[1] + temp_2 * challenger_binary_1v.glm$coef[2])/(1 +
    exp(challenger_binary_1v.glm$coef[1] + temp_2 * challenger_binary_1v.glm$coef[2]))
```

```
## (Intercept)
##   0.9996088
```

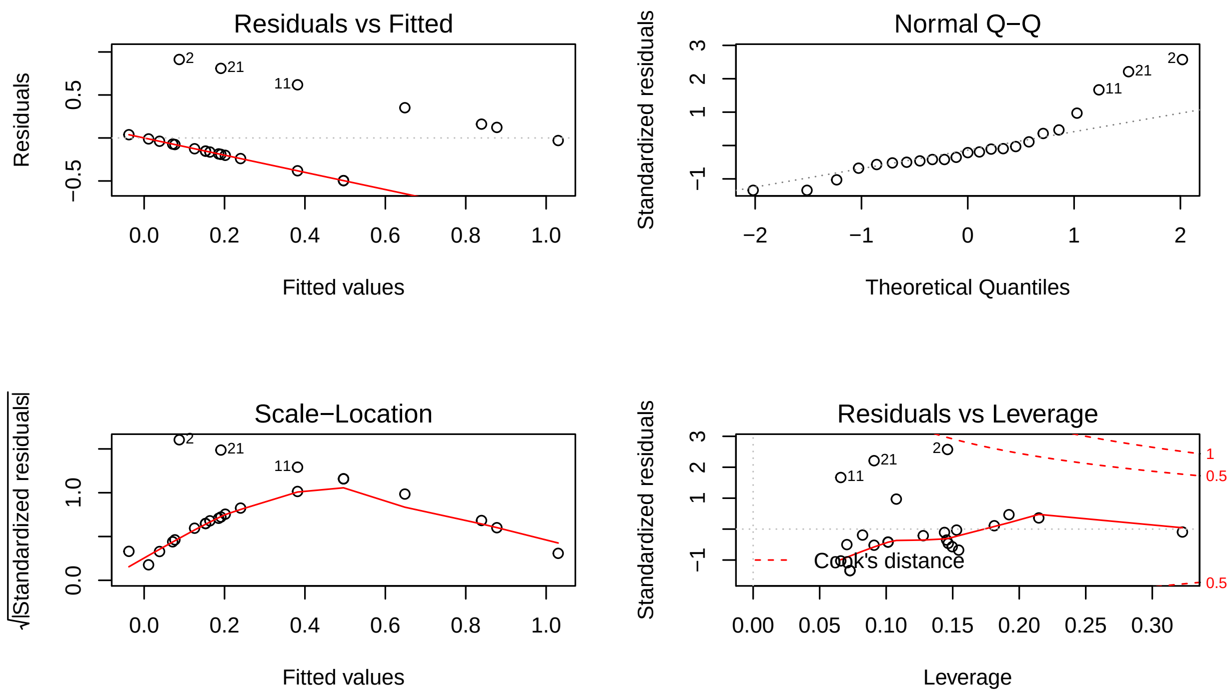
The estimated probability of a failure of an O-ring at the average temperature at the start of a challenger mission is around 24.83%, while at 31 degrees it is around 99.96%.

b. With the same set of explanatory variables in your final model, estimate a linear regression model. Explain the model results; conduct model diagnostic; and assess the validity of the model assumptions. Would you use the linear regression model or binary logistic regression in this case. Please explain.

```
# Fitting a linear model
linear_model <- lm(as.numeric(Incident) - 1 ~ Temp + Pressure,
    data = challenger)
summary(linear_model)
```

```
##
## Call:
## lm(formula = as.numeric(Incident) - 1 ~ Temp + Pressure, data = challenger)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.49600 -0.18816 -0.07650  0.08027  0.91267
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.658765   0.824297   3.225  0.00424 **
## Temp        -0.038136   0.011602  -3.287  0.00369 **
## Pressure     0.001962   0.001200   1.635  0.11779
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3837 on 20 degrees of freedom
## Multiple R-squared:  0.3952, Adjusted R-squared:  0.3347
## F-statistic: 6.534 on 2 and 20 DF,  p-value: 0.006549
```

```
par(mfrow = c(2, 2))
plot(linear_model)
```




```r
vif(linear_model)
```

```
##     Temp Pressure
## 1.001588 1.001588
```

We estimate a model with *Temp* and *Pressure*. We get:

$$\hat{P}(Incident) = 2.658 - 0.038 \times Temp + 0.002 \times Pressure$$

The summary of our binary linear regression model indicates that the temperature and intercept are statistically significant at 1%. This is similar to the logistic regression model results from earlier, with the pressure explanatory variable not showing a high level of significance. However, as we will note below, we will not be able to rely on the test statistic as the model assumptions are violated.

The Normal Q-Q plot shows that the data is not normally distributed given the divergence from the diagonal line in the tail quantiles. We note that we can also not rely on the Central Limit Theorem for the normality assumption, as the sample size is below 30. Temperature does not look necessarily normal. Independence is not given between observations and hence error terms. The number of failures is not normally distributed as well. We evaluate if heteroskedasticity is present by looking at the Residuals vs Fitted and Scale-Location plots, noting that the pattern of the red line is not flat. This indicates an inconsistent variance across our fitted values, thus the homoscedasticity assumption is also violated for our linear model. We check the variance inflation factors for our variables to see if multicollinearity is present, noting that they are both low, illustrating multicolinearity between the variables is not present.

In the case of using a binary linear regression model or a binary logistic regression model, one of the key factors is selecting an appropriate model based on the range of outcomes. In the case of linear regression, we could have outcomes that are negative or >1, but for logistic regression, the range is bounded to $[0, 1]$. Our outcome variable is also categorical, so linear regression would not be appropriate in this context, as the dependent variable is treated as continuous in linear regressions. Thus, we find that logistic regression is more appropriate given the nature of our dependent variable.

4. Conclusion

The question we investigated in this report concerns the risk of O-ring failures among space shuttle launches; namely, can observable factors such as launch temperature and combustion pressure help predict the probability of O-ring failures for space shuttles. Our question and analysis is based on the Dalal et al (1989) paper which studies this question. This analysis is important because it shows that statistically based risk management can inform decision making in high stakes environments. Space shuttle launches are a prime example of a high stakes environment, where failure is associated with severe consequences and loss of life, as evidenced by the Challenger tragedy. The original paper was intended to inform the risk of O-ring failure for the tragic Challenger shuttle crash - which was launched at a temperature (31F) over 20 degrees lower than previous launches in our data.

After examining the data as provided by Bilder and Loughin and conducting EDA, we evaluated candidate model specifications to predict O-ring 'failure' from Temperature and Pressure. We noted that failure can be interpreted as the failure of at least one O-ring (with a binary outcome) or number of O-ring failures (with binomial outcome and potential counts 0 to 6). We used the binary logistic regression model for several reasons. First, the binomial model would require each O-ring failure to be independent from each other. Given that a single space shuttle has 6 O-rings and there are many unobserved variables (such as production runs and installation factors), this struck us as implausible. We also assessed the trade-off associated with using the binary logistic regression model; namely, we lose information about how many O-rings failed per mission. As we have only two space missions with more than one O-ring failure in our dataset, we consider the loss of information to be small. These were two of the main factors among others that we chose a binary logistic regression model as our reference model in this analysis.

The main result of our analysis is that O-ring failure probability is related to temperature, where lower temperatures have a higher risk of at least one O-ring failure. The Challenger mission, which launched at 31F, is predicted in our model to have at least one O-ring failure with over 99% probability, and even at a lower Wald CI bound, probability was over 48%. Our result follows the result from Dalal et al. that the probability of an O-ring failure for the Challenger launch was high and that this result is also statistically significant.

In summary, this analysis illuminates that launch decisions can be better informed by understanding failure risk and setting risk tolerance levels on an informed basis. And most beneficially, this analysis has the potential to power decisions that avert disaster and save lives.