Space Shuttle Safety and the *Challenger*

Statistical Modeling for Critical Decisions

Aidan Jackson, Sandip Panesar, Devesh Khandelwal

Investigation of the 1986 Space Shuttle *Challenger* Accident

Summary

This exercise examines the Space Shuttle *Challenger*'s 1986 disaster using the previous safety record of the Space Shuttle Program. The probability of O-ring failures at various launch temperatures are modeled to make a prediction for the conditions on the morning of the disaster. It is found that the probability of a catastrophic failure is XX%, over XX times higher than when using the methods of the original estimate. The Rogers Commission, tasked with investigating the accident, noted that this underestimate occurred due to omitting launch data where no O-ring damage was found.[2] The original estimate procedure was used by Morton Thiokol, a contractor to NASA and producer of the O-ring involved component that failed.

This work was originally completed as part of the W271 Statistical Methods for Discrete Response, Time Series, and Panel Data course in the Master of Information and Data Science program at University of California, Berkeley. The exercise in turn was based on a previous 1989 paper on the same topic, originally published in the Journal of the American Statistical Association and included in this repo as a pdf.[1]

Background

On January 28th, 1986, the *Challenger* launched from Cape Canaveral, Florida, for what was the 25th orbital mission of the Space Shuttle program. 59 seconds after launch, a tracking camera recorded a burning plume coming out of the side of the right solid rocket booster (SRB). As the leaking fuel continued to burn, gradually other parts of the shuttle stack began to disintegrate. 73 seconds after launch, still visible in the sky, the shuttle exploded as different fuel sources ignited. This was particularly shocking both for being the first American astronaut deaths in a launch mission[1], as well as its public broadcasting as part of the inaugural Teacher in Space Project.

In the ensuing investigation, the cause of the accident was found to be O-ring failure in the field-joint of the same right SRB.[2] O-rings are doughnut-shaped and commonly made of rubber, and provide a flexible seal between mechanical components. In a typical car engine, O-rings have diameters on the scale of inches. Inside the Space Shuttles' SRBs, O-ring diameter was 37.5 feet.[1]

Like many other applications, the SRBs' O-rings sealed against gas to prevent it from escaping where mechanical joints connected. Every Shuttle launch required 2 SRBs, with each SRB containing 3 field-joints that used O-ring seals. For redundancy, each field-joint had both a primary and a secondary O-ring. This resulted in 6 field-joint O-rings total per SRB, or 12 field-joint O-rings total per launch.

Redundancy is needed because of the potential for an O-ring to fail. There are two mechanisms which could cause this, known as erosion and blow by. Erosion has the potential to occur during normal operation, where hot gases sealed against the O-ring gradually disintegrate it. On the SRBs, putty was added to provide a protective barrier against this mechanism. Blow by occurs when an O-ring already fails to provide a complete seal, allowing high velocity gas to move past the O-ring and damage it further.

---

[1]Apollo 1 resulted in 3 astronaut deaths during pre-launch training on the ground. Soyuz 1 and 11 resulted in a combined 4 cosmonaut deaths due to hardware failures during each launch mission.

Although the secondary O-rings were a backup should distress and failure occur in a primary O-ring, before the first ever Space Shuttle launch it was discovered that additional problems could also occur. This was due to field-joint design which sometimes caused unintended rotation among different components, including the O-rings. While the primary O-rings would engage a seal reliably even with this rotation, the secondary O-rings specifically could sometimes slip and leave permanent space for gas leakage. If this occurred during a launch where the primary O-ring failed due to the previous distresses, then it was known the result would be a "catastrophic failure".[1]

Being made of rubber-like material, there are a variety of conditions which may impact the ability of any O-ring to provide a complete seal and suffer distress. One which was investigated for the Space Shuttle program was temperature, where colder conditions can cause an O-ring to shrink and lose its elasticity. During the late 1970s and early 1980s, the SRB engine with the O-rings was qualified for temperatures as low as 40°F.

Without any additional statistics, when the Challenger started its Winter morning launch it was 32°F.

Exploratory Data Analysis

As with any dataset, understanding what information is available is a prerequisite to deciding on what value to extract from it. Although the size of this dataset is small, with only 23 observations from previous Shuttle launches, the choice of how to treat it is equally as important and influential to later analysis.

```
library(data.table)
library(ggplot2)
library(gridExtra)
library(dplyr)
library(mcprofile)

d <- read.csv("challenger.csv")
d <- data.table(d)

grid.table(head(d), rows = NULL)
```

| Flight	Temp	Pressure	O.ring	Number
1	66	50	0	6
2	70	50	1	6
3	69	50	0	6
4	68	50	0	6
5	67	50	0	6
6	72	50	0	6

With the data loaded, it can be seen that there are 5 variables and 23 observations. Between these, there are 0 missing values. In total there is a small amount of information available, which is to be

expected with rare events such as Space Shuttle launches. However, it provides an opportunity to more thoroughly examine all aspects of the dataset.

To begin, the meaning of each variable must be properly understood. `Flight` represents a generic index for each launch. `O.ring` represents the number of primary field-joint O-rings which experienced a "thermal distress" event for each flight. A thermal distress event was defined as when either the effects of erosion or blow by could be observed in the recovered SRBs post launch. Although every launch had 12 field-joint O-rings in total, only six of these are primary while the second six are secondary. This was to design for redundancy, even though mechanisms were found which could lead to the immediate failure of the secondary O-rings should the primary fail as mentioned in **Background**. As confirmation of the total, the variable `Number` represents the count of primary field-joint O-rings and is constant at six for all launches. `Temp` and `Pressure` represent the recorded field-joint temperature in Fahrenheit and pressure in psi respectively. `Temp` is recorded to the nearest degree, while `Pressure` is recorded to the nearest 50 psi (e.g. 50, 100, 150, etc). With a much coarser grain of pressure measurements, it may be less meaningful since less specific information is recorded.

Being the cause of the accident, the `O.ring` variable will be examined to start. In this analysis, unless stated otherwise "O-ring" will refer specifically to a primary field-joint O-ring.

```
ggplot(d, aes(O.ring)) +
  geom_histogram(binwidth = 1, fill = "black", color='black', alpha=0.8) +
  ggtitle("Figure 1. O-Ring Distribution") +
  xlab("Count of O-Ring Thermal Distress Events per Launch") +
  ylab("Count")
```


Figure 1. O–Ring Distribution
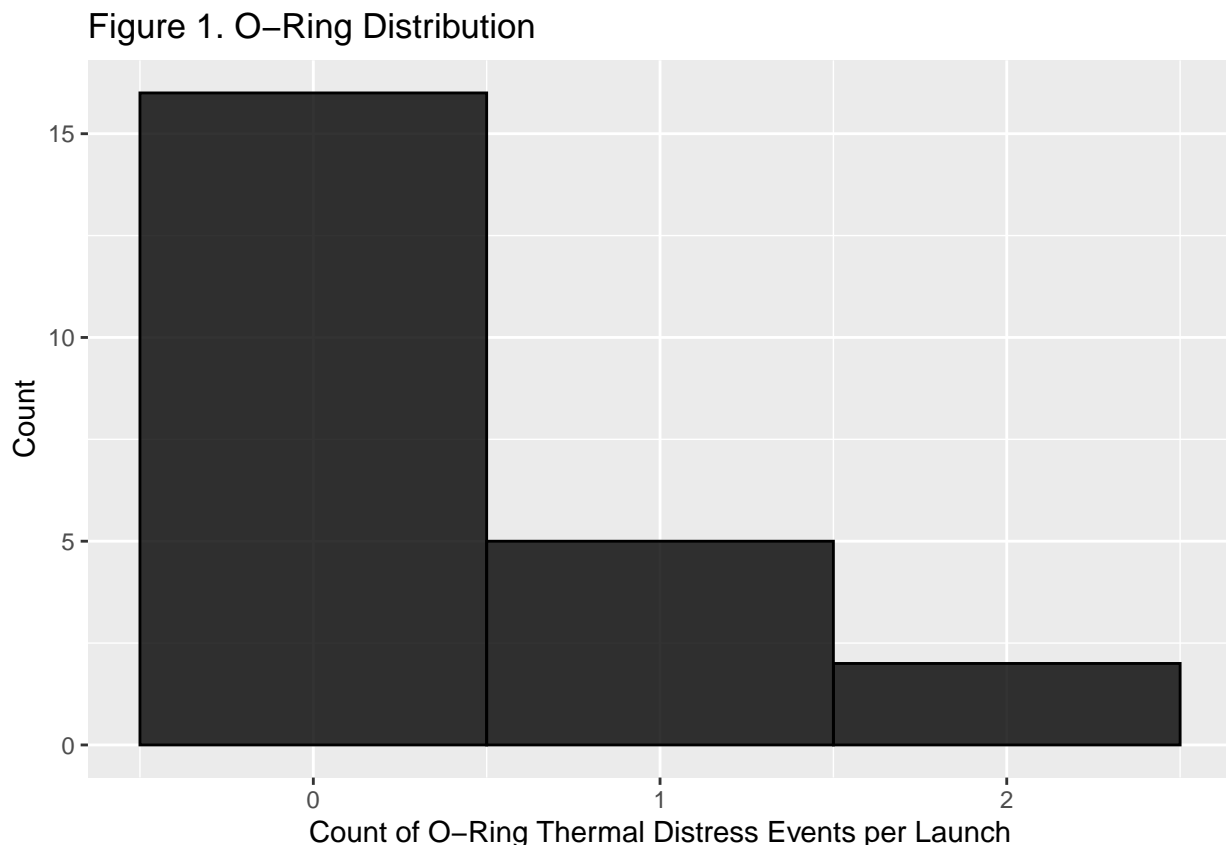
Shown above in **Figure 1**, the majority of launches had O-rings which experienced no thermal distress events. About 5 launches had a single O-ring experience thermal distress, while even fewer had two. None of the recorded launches had greater than two O-rings distress events. However, for a catastrophic failure to occur, all that was potentially needed would be a single primary O-ring failure to occur when its secondary backup had rotated and lost contact for a seal.

With the majority of launches not experiencing O-ring distress, those that experienced any distress at all can be grouped together for a more discrete analysis.

```
d$distress <- d$O.ring > 0

grid.table(tail(d), rows = NULL)
```

Flight	Temp	Pressure	O.ring	Number	distress
18	81	200	0	6	0
19	76	200	0	6	0
20	79	200	0	6	0
21	75	200	2	6	1
22	76	200	0	6	0
23	58	200	1	6	1

To better differentiate between these two situations, a new Boolean variable is created which indicates whether a launch had at least one O-ring distress event or not. With two potential predictors of these failures, `Temp` and `Pressure`, it will be important to understand how they relate to the distribution of the outcome.

```
ggplot(d, aes(factor(distress), Temp)) +
  geom_boxplot(aes(fill = factor(distress))) +
  theme(legend.position = "none") +
  geom_jitter() +
  ggtitle("Figure 2. Temperature vs. O-Ring Distress Event") +
  xlab("O-Ring Distress Event") +
  ylab("Temperature (°F)")
```

Figure 2. Temperature vs. O–Ring Distress Event

Demonstrated in **Figure 2**, O-Ring distress events generally occurred at lower launch temperatures. In fact, with so few total observations, it can be seen that over half of the launches with O-ring distress were at lower temperatures than the coldest launch with no O-ring distress There is a single outlier that is notable, however, where a distress event occurred at a launch temperature of ~75°F. This temperature would be greater than average even among launches which experienced no distress. Note that because of this point, along with generally having a wider distribution, there is greater temperature variance among launches which had distress events compared to those which did not.

The other environmental variable in the data set, `Pressure`, may also be examined to see if insightful relationships emerge.

```
ggplot(data = d, aes(Temp, O.ring, color = Pressure)) +
  geom_point() +
  ggtitle("Figure 3. O-Ring Failures vs. Environmental Conditions") +
  xlab("Temperature (°F)") +
  ylab("O-Ring Distress Events") +
  labs(size = "Pressure (psi)") +
  scale_color_gradient(low = "#56B1F7", high = "#132B43")
```

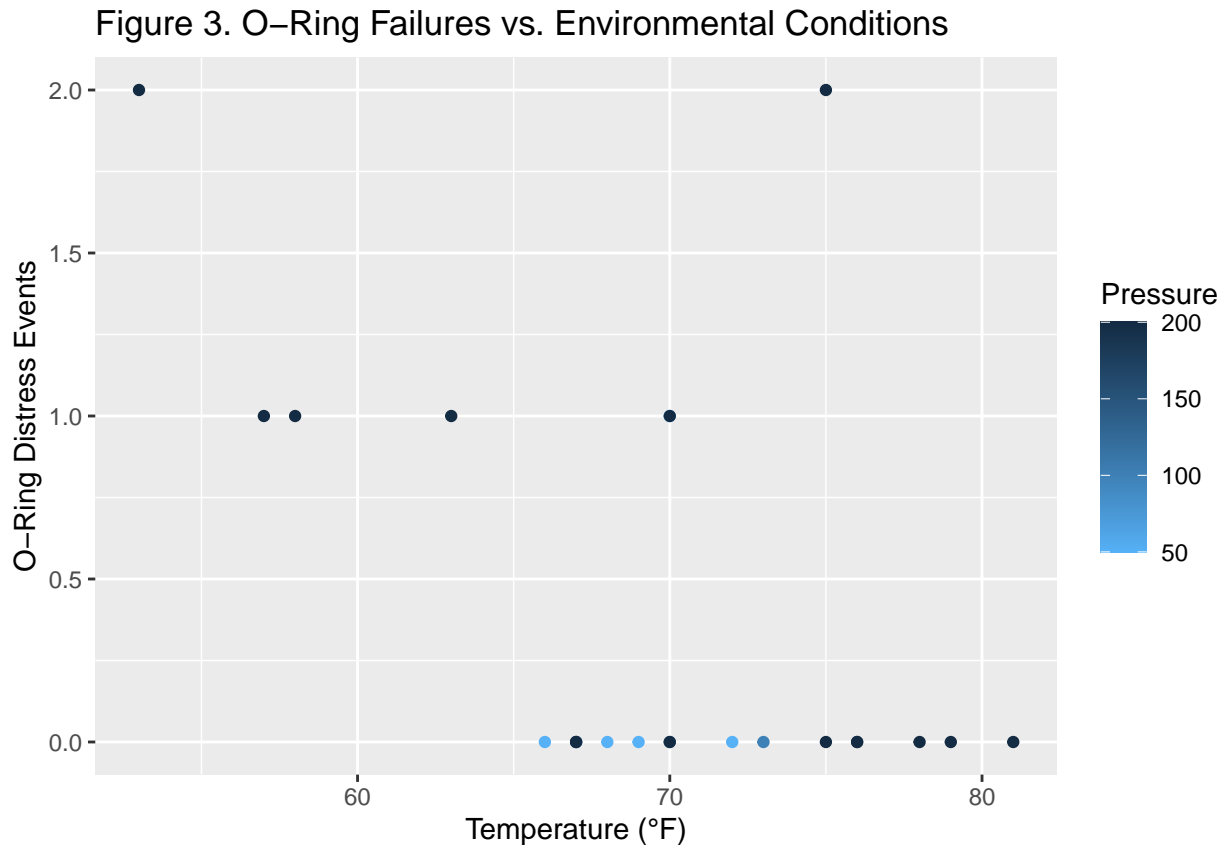Figure 3. O–Ring Failures vs. Environmental Conditions

Figure 3 shows that generally the launches with a greater number of O-ring distress events occurred at lower temperatures and higher pressures. This also fits with conceptual information known about the O-rings, such as lower temperatures causing them to shrink in size and provide a less complete seal against gases. Higher pressures could then also cause more erosion/blowby from a greater amount of gas in the same space in the system. While all of the distress events did occur at the highest pressure measurement of 200 psi, the general trend does not seem as strong as with temperature. For example, 200 psi pressure measurement was also the most common value of this variable generally. However, temperatures around the midpoint of the range ~65-75°F also had lower pressures, which may have helped them avoid any distress events.

The correlation between these variables can also be computed numerically for a more thorough understanding. Although pressure is binned into multiples of 50 psi, resembling a categorical or ordinal variable, the ratio of one pressure to another is still meaningful. For example, 200 psi is known to be four times as much pressure as 50 psi. A pressure of zero psi is also meaningful, meeting the definition for a ratio variable. Since theoretically a pressure can take on any value, even if only multiples of 50 are recorded, it will still be treated as a continuous variable. The count of O-rings which experienced distress meets the same definition of a ratio variable but is by nature discrete and not continuous. For example, 1.5 O-rings cannot experience distress. Therefore, Spearman's rank correlation coefficient will be used.

```
cat("Table 1. Correlation of Numerical Variables\n")
```

Table 1. Correlation of Numerical Variables

```
# generate correlation table without certain columns
cor(d[ , !c("Flight","Number","distress")], method = "spearman")
```

```
##                Temp   Pressure     O.ring
## Temp      1.0000000 0.1541326 -0.4294734
## Pressure  0.1541326 1.0000000  0.2775559
## O.ring   -0.4294734 0.2775559  1.0000000
```

Table 1 shows the correlation between the three numerical variables which varied in the data set over the course of the launches, being `Temp`, `Pressure`, and `O.ring`. As suspected in **Figure 3**, pressure had a positive correlation with O-ring distress incidents. This supports the previous observation that higher pressures were associated with both launches with and without distress events, but lower pressures almost always were on flights without any distress events.

Temperature had a stronger negative correlation with O-ring distress events. This continues to follow the trend shown in **Figures 2** and **3** where flights which experienced any distress were generally at colder temperatures. Finally, we can also see that temperature and pressure have a much smaller and slightly positive correlation with each other. This is in agreement with physical science, where gas enclosed in fixed volume will grow to higher pressures if the temperature is increased. Since the function of the O-ring is to provide a seal, i.e. a fixed volume, it would be reasonable that there is a slight positive correlation between the two.

Modeling

With the given context, there are different potential approaches to estimate the relationship between the environmental variables and O-ring distress. Given that each O-ring is discrete, a Poisson or Binomial distribution could be used depending on the scenario and modeling goals. The Poisson distribution describes processes which, among other things, occur over an interval at a constant average rate. This could be used to model the expected count of O-rings which fail for a series of Space Shuttle launches. However, the focus on the average over time is not particularly useful to estimate failure in a specific future launch. The Binomial distribution, on the other hand, describes processes which have a sequence of binary events, such as success or failure of multiple O-rings within a single Shuttle launch. This would be much more fitting for predicting a probability of failure on a single launch in the future, such as the *Challenger* at the start of its final mission.

For both the Poisson and Binomial distributions, an assumption of independence is required. This would require that each Space Shuttle launch is independent and identically distributed (IID) compared to all others. In general this should be true, although there may be minor reasons why this is violated. For example, over time Shuttle technology or launch procedures may have been improved based on prior experience. This would be equivalent to a distributional change over time. In addition, part of the reason SRBs were recovered were not just for observing O-ring damage, but also for refurbishment and potential re-use in later launches. Having sets of launches with the same SRBs could result in sets of correlated observations that are different from other flights. It could also be expected that re-used SRBs could in general cause greater O-ring distress than those being newly used. This would be a notable violation since it can involve the observed outcome of O-ring damage along with its relationship to environmental conditions. With either a Poisson or Binomial distribution, it must be assumed to be negligent. In the future, knowledge of which SRBs were re-used for which launches could be incorporated into modeling for a more accurate estimate.

Because of its better fit for the use case of predicting the probability of the *Challenger* disaster, the Binomial distribution will be used. Specifically, it will feature $n = 6$ IID O-rings on each flight, each with a probability p of experiencing distress. In addition to each Shuttle launch being IID, each of the six O-rings for any given flight must be IID as well. If the O-rings were not independent, the damage of a given O-ring would not only be a function of the environmental variables but also of the damage of other O-rings. Consequently, there would not be just one relationship between the O-ring damage and the predictors, but multiple relationships depending on which O-ring was under consideration.

Just like with the IID Shuttle launches, this is avoided with the assumption of independence however it may not be completely valid. For example, the six O-rings are spread out across two SRBs for each launch. It could be expected that there are small differences from between specific SRBs with respect to the stress the O-rings experience. For example, if an O-ring in one SRB had minor manufacturing differences that put more stress on its O-rings then these observations would be correlated. The three O-rings in the other SRB, however, may have their own correlation based on its own unique qualities. Therefore, in reality there is likely some relationship between the O-rings within a single SRB that is ignored for the sake of the model.

(b) Estimate the logistic regression model using the explanatory variables in a linear form.

There are two explanatory variables of interest, `Temp` and `Pressure`. First `Temp` is fit in a standalone model, and then another with both `Temp` and `Pressure`.

```
# Null hypothesis- coefficient of pressure is 0
model_h0 <-  glm(formula = distress ~ Temp, data = d ,
                family = binomial(link = logit))


# Alt hypothesis- coefficient of pressure is not 0
model_h1 <- glm(formula = distress ~ Temp + Pressure, data = d ,
                family = binomial(link = logit))
```

(c) Perform LRTs to judge the importance of the explanatory variables in the model.

In Part B, two models were fit to test the hypothesis:

$$H_0 : \beta_2 = 0$$

$$H_a : \beta_2 \neq 0$$

(where β_2 is the model coefficient of `Pressure`). Note that the statistical significance of `Temp` is demonstrated in Section 3-A.

A Likelihood Ratio (LR) test will be performed using the function *anova*. In general an LR test can be mathematically expressed as:

$$-2log(\Lambda) = -2log(\frac{L(\hat{\beta}^{(0)}|y_1, \ldots, y_n)}{L(\hat{\beta}^{(a)}|y_1, \ldots, y_n)}) = -2\sum y_i log\left(\frac{\hat{\pi}_i^{(0)}}{\hat{\pi}_i^{(a)}}\right) + (1 - y_i)log\left(\frac{1 - \hat{\pi}_i^{(0)}}{1 - \hat{\pi}_i^{(a)}}\right)$$

Where:

- $\hat{\pi}_i^{(0)}$ is the estimated probability of success under H_0

- $\hat{\pi}_i^{(a)}$ is the estimated probability of success under H_a

Code for the anova function is:

```
anova_obj <- anova(model_h0 , model_h1, test = "Chisq")
anova_obj
```

```
## Analysis of Deviance Table
##
## Model 1: distress ~ Temp
## Model 2: distress ~ Temp + Pressure
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        21     20.315
## 2        20     18.782  1   1.5331   0.2156
```

```
beta_est <- anova_obj$`Pr(>Chi)`[2]
```

From analyzing the anova outputs, we see that the probability of obtaining a non-zero value of co-efficient of `Pressure` is 0.2156479, hence there is not enough evidence to reject the null-hypothesis.

(d) The authors chose to remove Pressure from the model based on the LRTs. Based on your results, discuss why you think this was done. Are there any potential problems with removing this variable?

In the previous LR test, there was not sufficient evidence to reject the null hypothesis, and thus the `Pressure` variable was found to be insignificant. This agrees with the authors' choice. To better understand this result, the confidence interval (CI) can also be examined. This helps to understand the uncertainty in this result with such a small number of samples in the data set.

```
pressure_ci <- confint(object = model_h1 , parm = "Pressure", level = .95)
```

It can be seen that the 95% Wald CI is between -0.0057514 and 0.0322629. As zero lies within the CI, there is additional evidence that the `Pressure` variable is not statistically significant from both Wald and the previous LR test. There is a failure to reject the null-hypothesis, which states that $\beta_{Pressure} = 0$.

To see how this could effect the odds of a launch failure, and cause a potential problem in the model, consider a 10 psi increase in `Pressure` with all else constant.

```
exp_pressure_ci <- exp(10*pressure_ci)
```

With every 10 psi increase in `Pressure` the odds of a launch failure changes between 0.9441084 and 1.3807523 times at the 95% confidence level. This re-demonstrates that `Pressure` is not significant, since the OR contains 1, but it does show that the majority of the predicted OR is greater than

one. An OR greater than one would indicate that an increase in the variable would lead to greater odds of the outcome event occuring. As was seen before, there did appear to be a slight positive correlation between higher pressures and O-ring distress, which likely led to most of the OR being above 1 in the interval. In addition, `Pressure` is a physical phenomena related to `Temperature` in the system, and only considering one and not the other may be taking a unnecessary risk. However, due to being statistically insignificant, there is not evidence that the model would perform better with its inclusion and so it will be dropped.

Part 3 (35 points)

Answer the following from Question 5 of Bilder and Loughin Section 2.4 Exercises (page 129-130):

Continuing Exercise 4, consider the simplified model $logit(\pi) = \beta_0 + \beta_1 Temp$, where π is the probability of an O-ring failure. Complete the following:

  (a) Estimate the model.

Above, we demonstrated that a model including the `Pressure` variable did not offer significantly more information compared to a model with only temperature as the dependent variable.

```
# Repeat with binomial model
# binomial_model <- glm(formula = distress/Number ~ Temp, data = d , family = binomial(link =
# This is the binary model
fail_model <- glm(formula = distress ~ Temp, data = d ,
                  family = binomial(link = logit))
se <- sqrt(diag(vcov(fail_model)))
```

In this model, an increase in the temperature by 1 degree Fahrenheit will cause a -0.232 decrease in the log-odds of an O-ring failure event. This coefficient value is associated with a standard error of 0.108, which is significant at the 95% level. This demonstrates that `Temp` is a significant variable for predicting the outcome.

  (b) Construct two plots: (1) π vs. Temp and (2) Expected number of failures vs. Temp. Use a temperature range of 31° to 81° on the x-axis even though the minimum temperature in the data set was 53°.

  (c) Include the 95% Wald confidence interval bands for π on the plot. Why are the bands much wider for lower temperatures than for higher temperatures?

First, we construct a data frame of X-values (i.e. temperature in degrees Fahrenheit), that will be used to predict π. Then, we feed this data frame into the model using the predict function, with the `response` function specified and `se=TRUE` . This returns a corresponding vector of probabilities for failure based upon the existing `fail_model` which was previously specified:

```
# Create a dataframe of new temperatures
temps <- data.frame(seq(31,81,1))
colnames(temps) = 'Temp'
```

11

```r
# Create vector of predictions using model_h0 (i.e. the model without pressure)
temp_prob <- predict(object=fail_model, newdata=temps, type='response', se=TRUE)

# Pi values
temp_pis <- temp_prob$fit
# SEs
temp_ses <- temp_prob$se.fit

alpha <- .05

#2.5th %ile
lower_ci <- temp_pis - qnorm(p = 1-alpha/2)*temp_prob$se
#97.5th %ile
upper_ci <- temp_pis + qnorm(p = 1-alpha/2)*temp_prob$se

final_temp <- data.frame(temps$Temp, temp_pis, lower_ci, upper_ci)
colnames(final_temp) <- c('temp', 'pi', 'lwr', 'upr')

ggplot(data=final_temp, aes(x=temp)) +
  geom_line(aes(y=pi)) +
  geom_line(aes(y=lwr), color='red', linetype='dotted') +
  geom_line(aes(y=upr), color='darkgreen', linetype='dotted') +
  ylim(0,1) +
  ggtitle("Figure 4. Probability of O.Ring Failure by Temperature") +
  ylab(expression(pi)) +
  xlab("Temp (F°)") +
  geom_hline(yintercept=1.0, linetype='dotted') +
  geom_hline(yintercept=0.0, linetype='dotted') +
  geom_point(data = d, aes(x=Temp, y= ifelse(distress==TRUE, 1, 0) ),
             shape = 17) +
  geom_text(aes(x = 58, y = 0.27), label = "2.5th %ile C.I.", color='red') +
  geom_text(aes(x = 78, y = 0.30), label = "97.5th %ile C.I.", color='darkgreen')
```
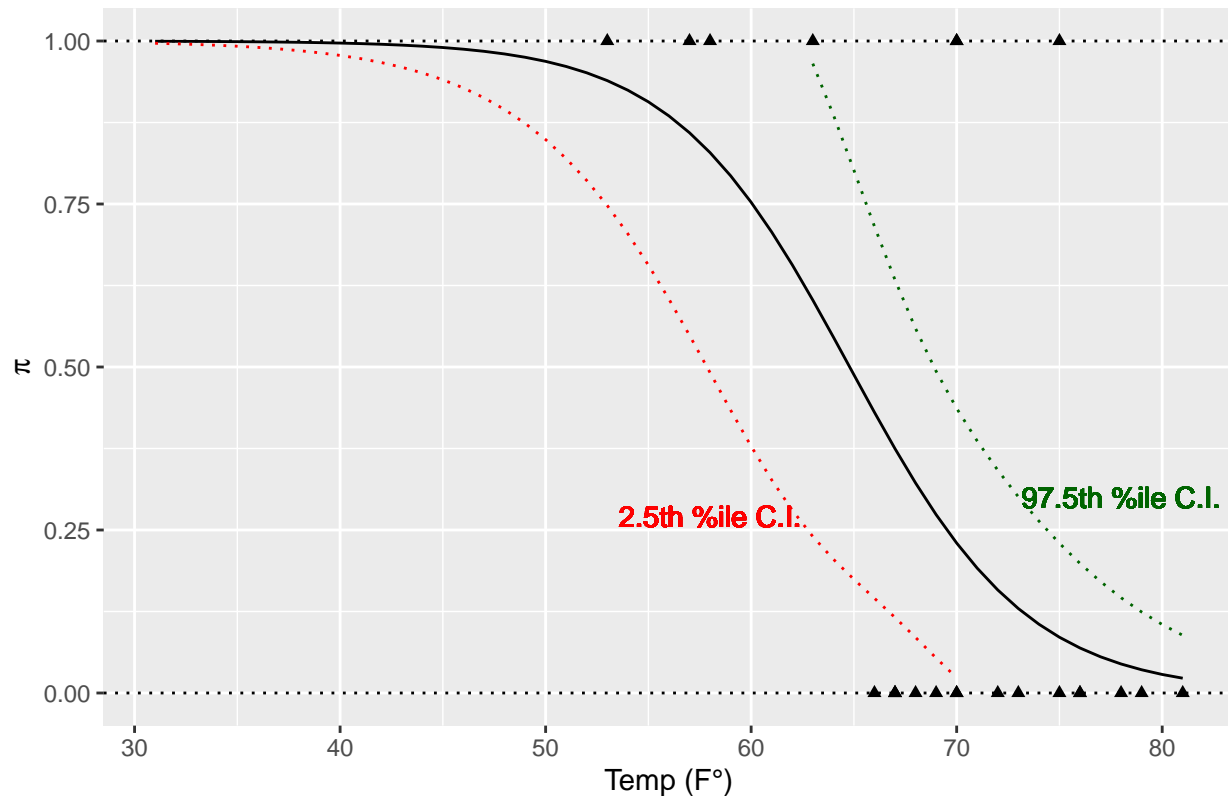
Figure 4. Probability of O.Ring Failure by Temperature

Shown above in *Figure 4*, the estimated probability of any O-ring failure is close to 1 at low temperatures, including the temperature at which the Challenger launched at.

From *Figure 4* we also observe that the width of the confidence interval (CI) increases gradually until ~55°F before decreasing. This is due to the fact that there are no ground truth observations below 50°F, and that the data is an approximation of the true distribution. Hence, there is a greater amount of sampling variance in $\hat{\pi}$ at these low temperatures.

```
ggplot(final_temp, aes(pi)) +
  geom_histogram(binwidth = .1, color='black', fill = "black", alpha=0.5) +
  xlim(0,1) +
  ylim(0,10) +
  ggtitle("Figure 5. Count Distribution of Predicted Probabilities") +
  xlab("Predicted Probability") +
  ylab("Count")
```

```
## Warning: Removed 2 rows containing missing values (geom_bar).
```

Figure 5. Count Distribution of Predicted Probabilities
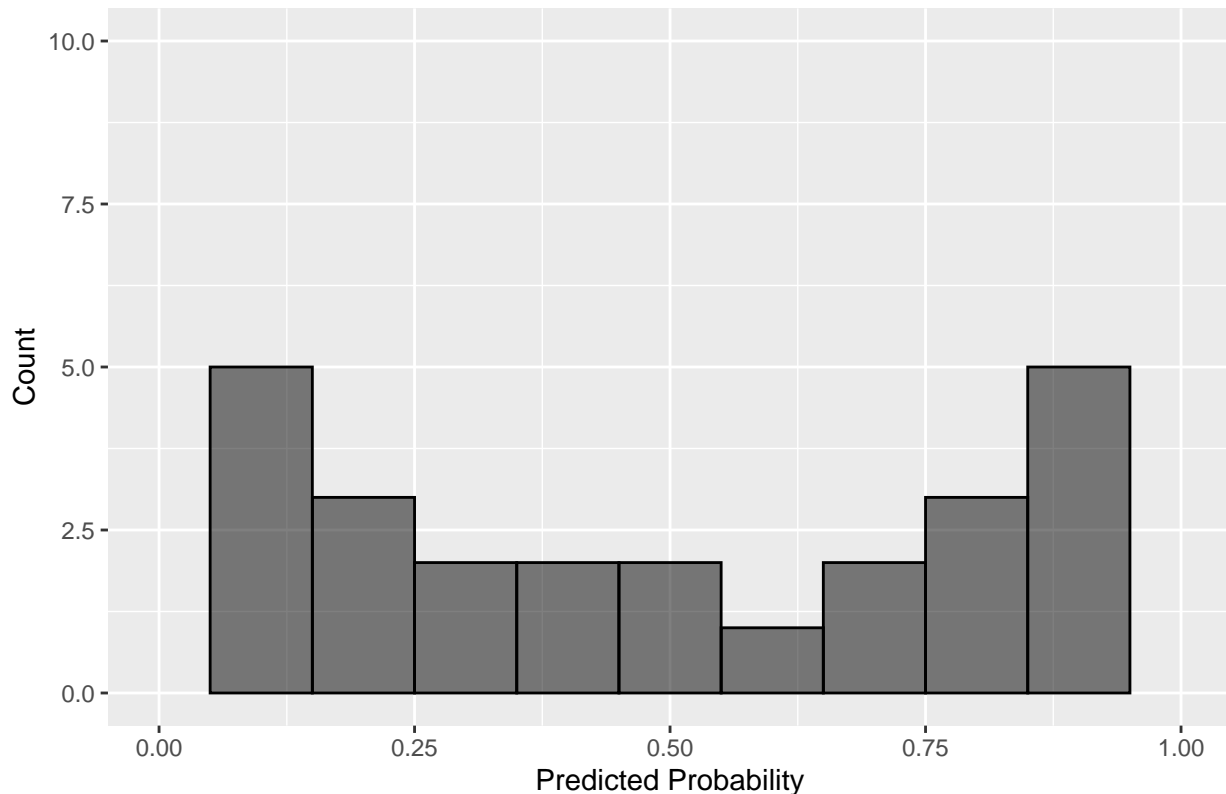
Moreover, the Wald interval assumes the outcome variable follows a normal distribution and has a large sample size, which is not true and demonstrated by *Figure 5* above. Shown in *Figure 5*, $\hat{\pi}$ has fewer observations towards the center of its range, with more on the edges. This is almost the opposite of a standard normal distribution, where the median value is the most often observed. Hence, the resulting interval is broader in general.

Next, the expected number of failures as a function of temperature can be calculated.

```
# This function calculates the maximum number of failures at a specified probability of a sing
get_ring_failure <- function(est_prob){
  # This function is from the paper Dalal et al.
  mu = 6.0*est_prob
  return(mu)
}
```

```
# Now calculate expected number of incidents

# Create vector of predictions using model_h0 (i.e. the model without pressure), also specify
temp_exps <- predict(object=fail_model, newdata=temps, type='response', se=TRUE)

# Pi values
pi_hat <- temp_exps$fit

# SEs
```

```
pi_hat_ses <- temp_exps$se.fit

#calculating failure
failure_hat <- apply(X = data.frame(pi_hat), MARGIN = 1, FUN = get_ring_failure)

#creating a data frame
final_temp_exp <- data.frame(temps$Temp, pi_hat, failure_hat)
colnames(final_temp_exp) <- c('Temp', 'pi_hat', 'failure_hat')

#creating a subset of the original df to only retain relevant cols:
d_trunc <- d[,c("Temp","O.ring")]

#merging the data frames to create a final plot:
final_temp_exp = merge(x = final_temp_exp, y = d_trunc, by = "Temp",
                                    all.x = TRUE)

 ggplot() +
  geom_line(data=final_temp_exp, aes(x=Temp, y= failure_hat)) +
  ggtitle("Figure 6. Expected Number of O-Ring Incidents by Temperature") +
  ylab("No. Incidents") + xlab("Temp (F°)") +
   geom_point(data = final_temp_exp , aes(x=Temp, y= O.ring ), shape = 17) +
   labs(colour = "final_temp_exp$O.ring")
```
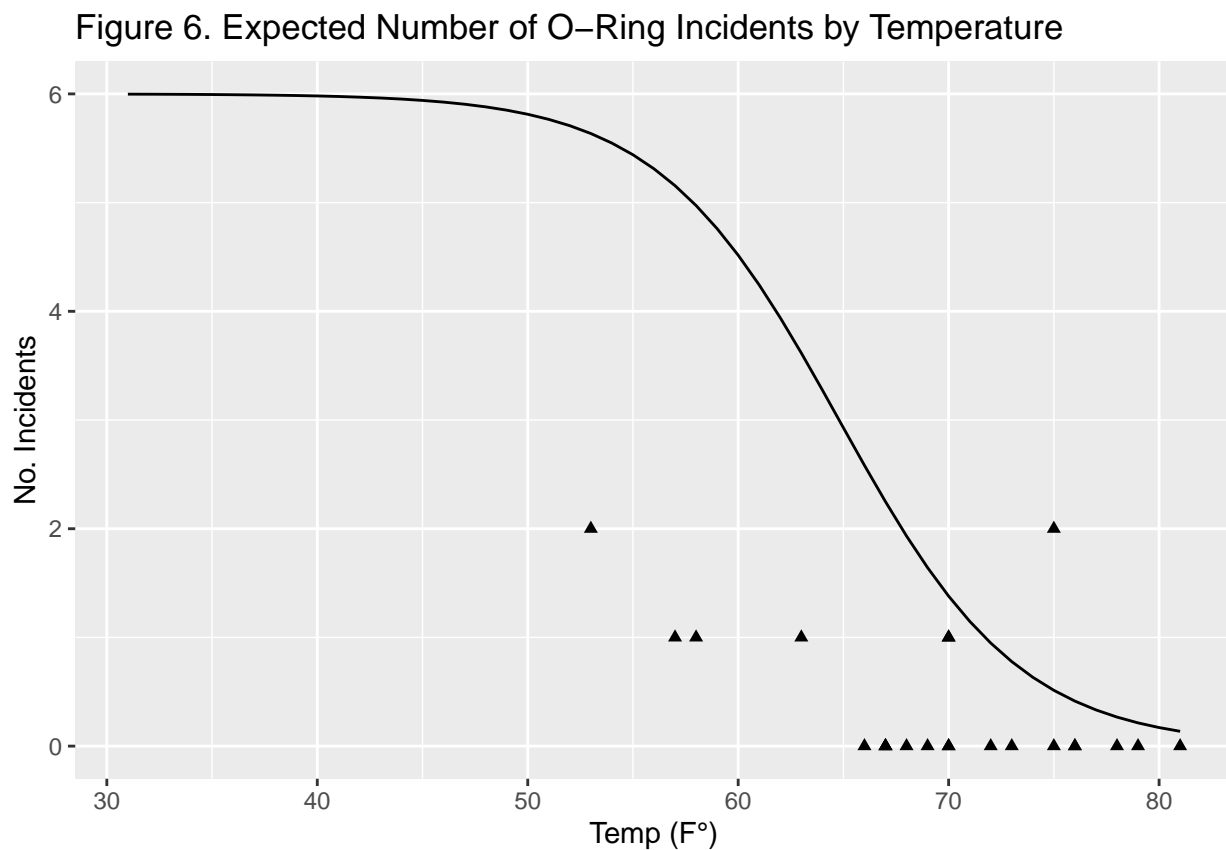

Figure 6. Expected Number of O−Ring Incidents by Temperature

Figure 6 above shows that it is expected that more O-ring failures than previously observed would have occurred at the Challenger's launch temperature. This should have caused serious concern for launching at the temperature of that day had this analysis been carried out ahead of time. The only point which does not fit this curve well is the observation with two O-ring failures at a temperature of about ~75°F.

(d) The temperature was 31° at launch for the Challenger in 1986. Estimate the probability of an O-ring failure using this temperature, and compute a corresponding confidence interval. Discuss what assumptions need to be made in order to apply the inference procedures.

We've already calculated the Wald intervals in the previous section. As shown before, a Wald test relies on following assumptions:

1. The underlying data is normally distributed.
2. We have a large sample size.

As we have seen in the previous subsection, $\hat{\pi}$ is not normally distributed and there is a small sample size. Therefore, in this section we'll look at estimating the CI with the Likelihood Ratio Test (LRT) using the model that we previously fitted. Unlike the Wald interval, the LRT assumes that the the the test statistic has a `chi.sq` distribution with 1 degree of freedom.

```
beta_0 <- fail_model$coefficients[1]
beta_1 <- fail_model$coefficients[2]

# Create a K-matrix for temp = 31 degrees

K <- matrix(data = c(1,31) , nrow = 1, ncol = 2)

# Calc -2log(lambda)
lc <- mcprofile(object=fail_model, CM=K)
ci_logit_profile <- confint(object=lc, level=0.95)

interval <- exp(ci_logit_profile$confint)/(1 + exp(ci_logit_profile$confint))

wald_int_profile <- wald(object=lc)

wald_ci_exp <- confint(wald_int_profile , level = .95)

wald_lower_exp <- wald_ci_exp$confint$lower
wald_upper_exp <- wald_ci_exp$confint$upper

wald_interval_lower <- exp(wald_lower_exp)/(1 + exp(wald_lower_exp))
wald_interval_upper <- exp(wald_upper_exp)/(1 + exp(wald_upper_exp))
```

For the LRT test and at 31°F, the odds of any `O.ring` failure has a 95% confidence interval of 0.8036982 and 1. The Wald test under the same conditions has the 95% confidence interval of any `O.ring` failure between 0.4816106 and 0.9999999. This demonstrates that the LRT is more conservative than the Wald and should be preferred.

(e) Rather than using Wald or profile LR intervals for the probability of failure, Dalal et al. (1989) use a parametric bootstrap to compute intervals. Their process was to (1) simulate a large number of data sets (n = 23 for each) from the estimated model of Temp; (2) estimate new models for each data set, say and (3) compute at a specific temperature of interest. The authors used the 0.05 and 0.95 observed quantiles from the simulated distribution as their 90% confidence interval limits. Using the parametric bootstrap, compute 90% confidence intervals separately at temperatures of 31° and 72°.27

First, the calculations will be set up.

```
bs_data_1 <- data.frame(c(31))
colnames(bs_data_1) = 'Temp'

bs_data_2 <- data.frame(c(72.27))
colnames(bs_data_2) = 'Temp'

vector_pi <- NA
vector_lower_ci <- NA
vector_upper_ci <- NA

vector_pi2 <- NA
vector_lower_ci2 <- NA
vector_upper_ci2 <- NA

suppressWarnings(for(i in 1:1000) {

  # generating data from repeated sampling, only temperature will be used
  bootstrap_temps <- sample_n(d, size=23, replace=TRUE)$Temp

  # make predictions on data with original model
  bootstrap_outcomes <- predict(object = fail_model, data = bootstrap_temps,
                      type = "response", family = binomial(link = "logit"))

  # put bootstrap data together for fitting new model
  bootstrap_d <- data.frame("Temp" = bootstrap_temps,
                      "distress" = bootstrap_outcomes)

  # fit new model using bootstrapped data
  model <- glm(formula = distress ~ Temp,
              data = bootstrap_d, family = binomial(link = logit))

  # make predictions using bootstrap model at temperatures of interest
  prediction_1 <- predict(object=model,
                      newdata=bs_data_1, type='response', se=TRUE)
  vector_pi[i] <- prediction_1$fit
  #note that we are supposed to do a 90% CI
  vector_upper_ci[i] <- prediction_1$fit + 1.645*prediction_1$se.fit
```

```r
    vector_lower_ci[i] <- prediction_1$fit - 1.645*prediction_1$se.fit
    #doing the same for 72 degrees
    prediction_2 <- predict(object=model,
                            newdata=bs_data_2, type='response', se=TRUE)
    vector_pi2[i] <- prediction_2$fit
    vector_upper_ci2[i] <- prediction_2$fit + 1.645*prediction_2$se.fit
    vector_lower_ci2[i] <- prediction_2$fit - 1.645*prediction_2$se.fit
})

temp_30_pis <- data.frame(data.frame(vector_pi),
                          data.frame(vector_lower_ci),
                          data.frame(vector_upper_ci))
colnames(temp_30_pis) = c('pi','lwr', 'upr')

temp_72_pis <-data.frame(data.frame(vector_pi2),
                          data.frame(vector_lower_ci2),
                          data.frame(vector_upper_ci2))
colnames(temp_72_pis) = c('pi','lwr', 'upr')


#calculating the standard deviations

pi_hat30_std_dev <- sd(temp_30_pis$pi)
pi_hat72_std_dev <- sd(temp_72_pis$pi)


#estimated mean
mu_hat_30 <- mean(temp_30_pis$pi)
mu_hat_72 <- mean(temp_72_pis$pi)

alpha <- .1
ci_30 <- mu_hat_30 + qnorm(p = c(alpha/2,1-alpha/2))*pi_hat30_std_dev
ci_72 <- mu_hat_72 + qnorm(p = c(alpha/2,1-alpha/2))*pi_hat72_std_dev
```

Next, the figures will be created.

```r
# 30 degrees
plt_30 <- ggplot(data = temp_30_pis) +
aes(x = pi) +
geom_histogram(binwidth=0.01, fill="blue", color="blue", alpha=0.3) +
geom_vline(xintercept = ci_30 , color = 'darkorange') +
labs(title = "Figure 7. Histogram of Bootstrapped Probabilities of Failure at 31°F",
     caption = ("Vertical lines denote the 90% CI")) +
  ylab("Count, Logarithm") +
  xlab(expression(pi)) +
  scale_y_log10()

#72 degrees
plt_72 <- ggplot(data = temp_72_pis) +
```

```
aes(x = pi) +
geom_histogram(binwidth=0.01, fill="purple", color="purple", alpha=0.3) +
geom_vline(xintercept = ci_72 , color = 'darkorange') +
labs(title =
        "Figure 8. Histogram of Bootstrapped Probabilities of Failure at 72.27°F",
    caption = ("Vertical lines denote the 90% CI")) +
    ylab("Count") +
  xlab(expression(pi))

plt_30
```
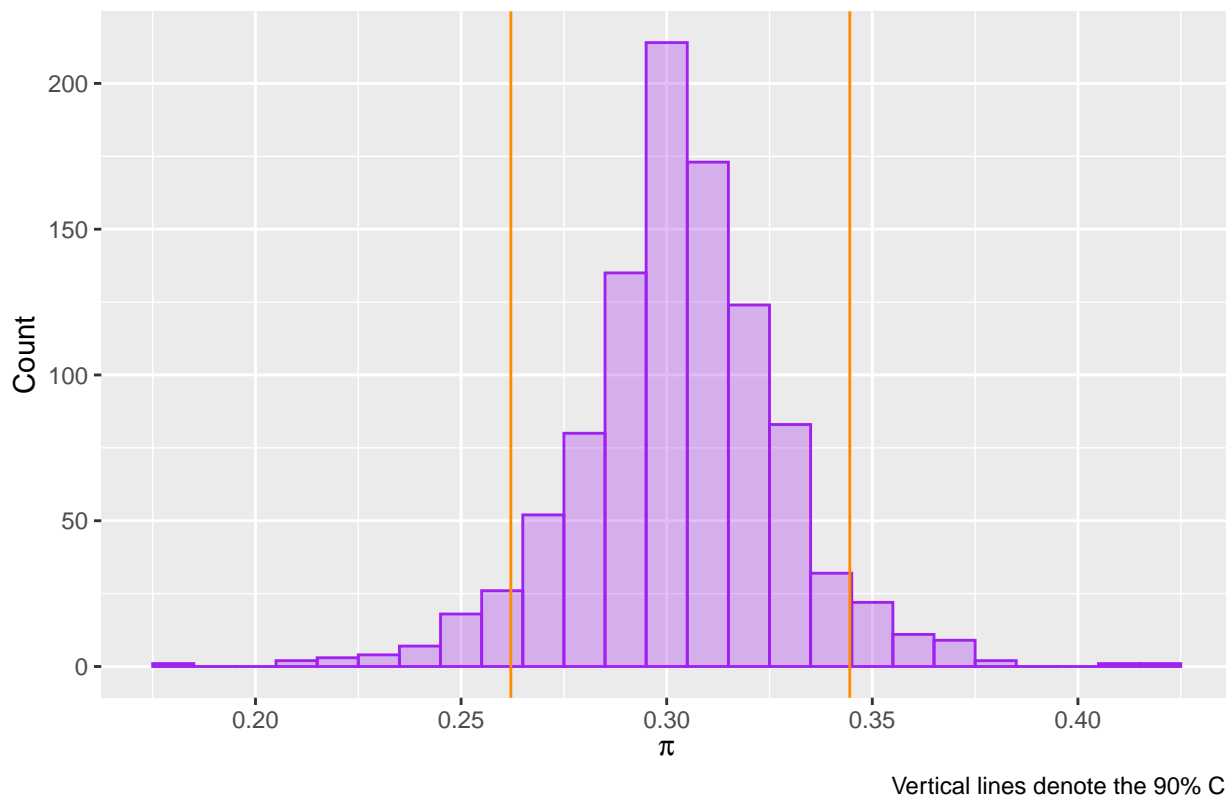
Warning: Transformation introduced infinite values in continuous y-axis

Warning: Removed 7 rows containing missing values (geom_bar).
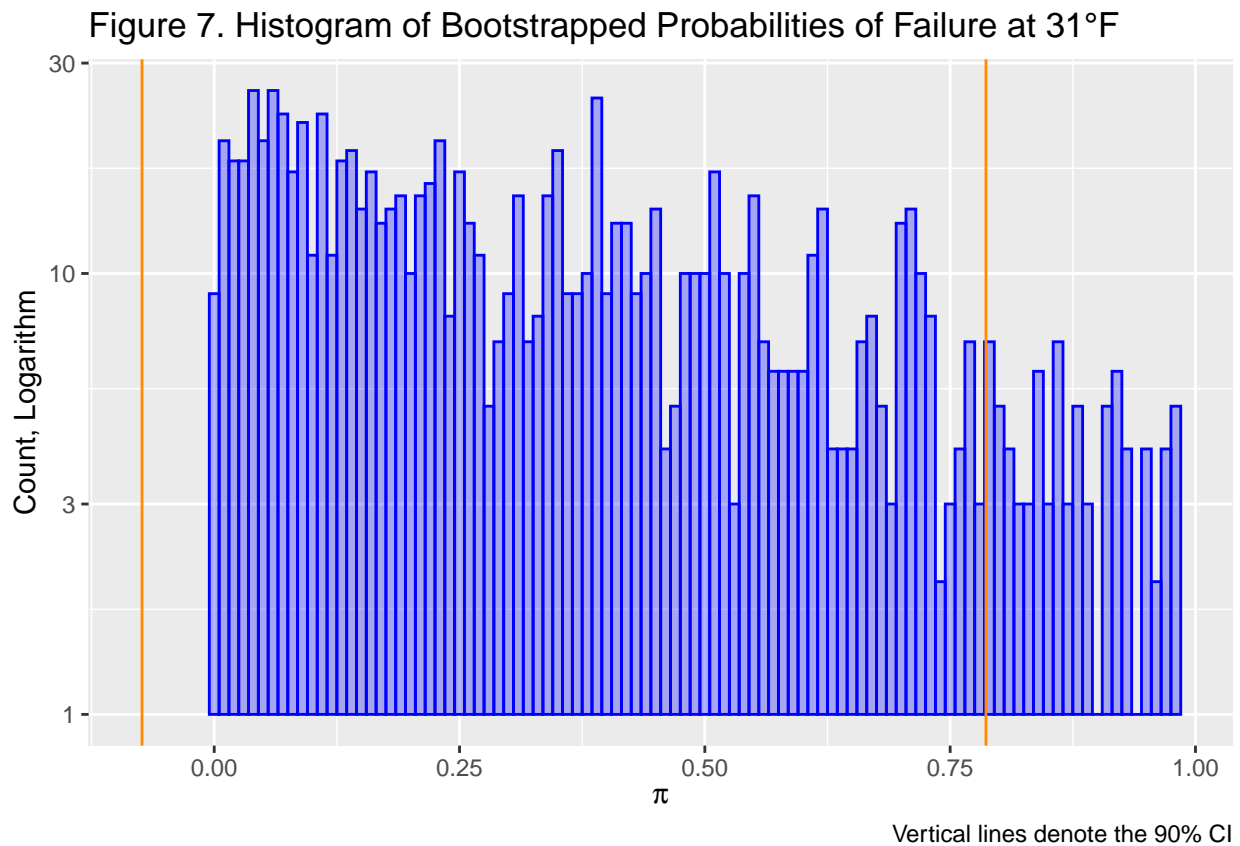
Figure 7. Histogram of Bootstrapped Probabilities of Failure at 31°F

```
plt_72
```

Figure 8. Histogram of Bootstrapped Probabilities of Failure at 72.27°F

Vertical lines denote the 90% CI

We can visualize a histogram of the results of the bootstrap (100,000 trials each) for 31°F in **Figure 7** and for 72°F in **Figure 8**. Using data in these figures, we can calculate that:

- The probability of an `O.ring` failure at 30°F are between -0.0735541 and 0.7865244 (with a 90% C.I.)

- The probability of an `O.ring` failure at 72.27°F are between 0.2620987 and 0.3445246 (with a 90% C.I.)

While these probabilities may be calculated to sometimes be less than one or greater than zero, it is known their true bounds are between [0, 1].

(f) Determine if a quadratic term is needed in the model for the temperature.

We'll create another glm model with a quadratic term and then compare the two models with an anova test as below:

```
fail_model_quad <- glm(formula = distress ~ Temp + I(Temp^2), data = d,
                       family = binomial(link = logit))
anova_mod <- anova(fail_model, fail_model_quad , test = "Chisq")

p_val<- anova_mod$`Pr(>Chi)`[2]
```

20

The p-value from the anova test is 0.3357766, which is fairly large and greater than 0.05. Thus, we conclude that there is not enough evidence that the quadratic model is an improvement over the the linear model.

Part 4 (10 points)

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

```
linear_model <- glm(distress ~ Temp, data = d, family = gaussian())
```

Shown above, the linear model also finds that temperature is significant for the presence of a O-ring distress event. Like the previous models there is an inverse relationship between the two variables, where the likelihood of a distress event increases as the temperature decreases. Specifically, with all else constant, increasing the temperature by 1°F would decrease the probability of a distress event by ~3.74%.

Even though a similar result is found, this model should not be used instead of the binary logistic regression model. This is because the linear model does not have bounds on the output, so that sometimes a probability greater than one or less than zero may be predicted. In this case, a probability of one is reached at about 50.96°F while a probability of zero is reached at about 77.71°F. Given that there were recorded launches both above and below these temperatures, this is an undesirable property of the model.

In addition, the linear model assumes homoscedasticity, or that the variance in the model's residuals is constant both over the range of the output as well as the domain of its inputs.

```
linear_model %>%
  ggplot(aes(x = linear_model$fitted.values,
             y = sqrt(abs(linear_model$residuals/sd(linear_model$residuals))))) +
  geom_point() +
  stat_smooth(color="red", se=FALSE) +
  labs(
    title = "Figure 9. Predicted Probabilities vs. Standardized Residuals",
    x = "Predicted Values",
    y = "sqrt(|Standardized Residuals|)")
```

```
## 'geom_smooth()' using method = 'loess' and formula 'y ~ x'
```

21

Figure 9. Predicted Probabilities vs. Standardized Residuals
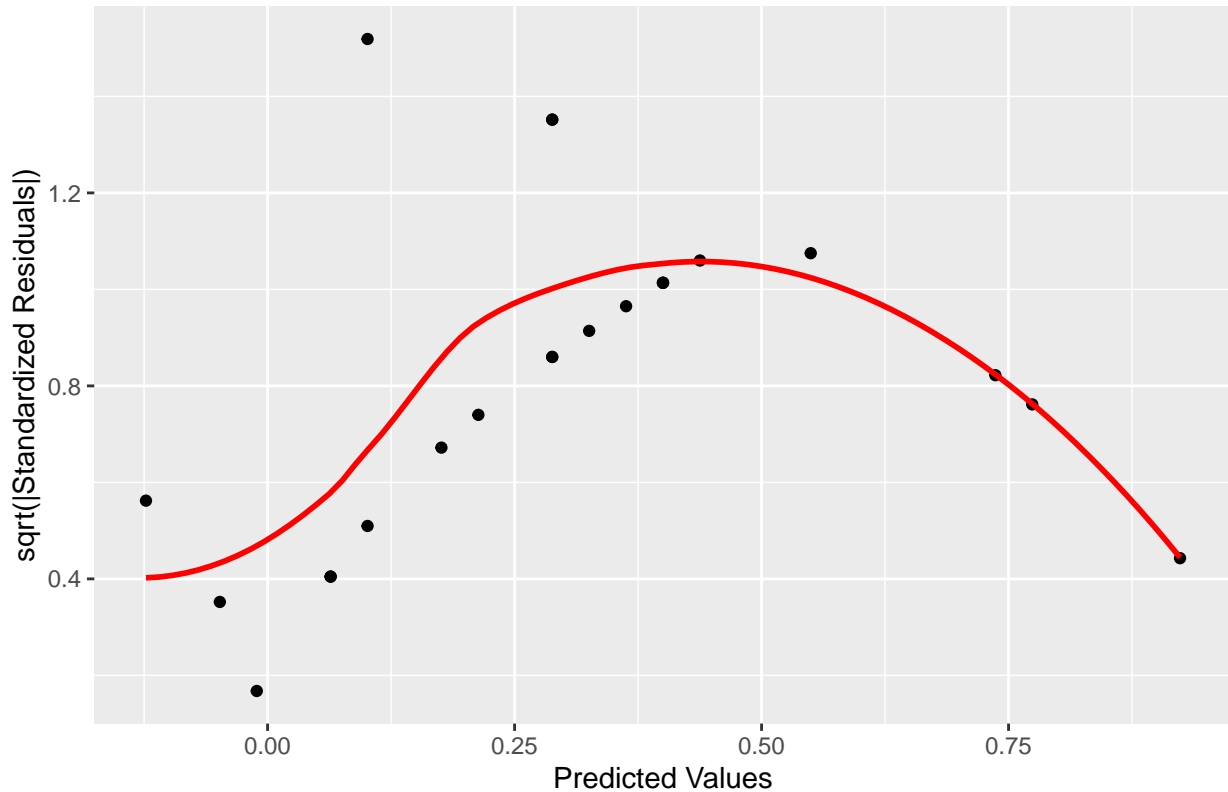
Shown in *Figure 9*, the standardized residuals are not constant over the range of the output variable in the model. This does not meet the homoscedasticity assumption, because the output variable is itself a function of the input variables. This result was expected, however, as it can be derived that the variance in a random variable from a binomial distribution is a function of that variable itself, or $Var(\pi) \propto \pi * (1 - \pi)$. It can be further seen that the maximum variance is observed around $\pi \approx 0.5$, which follows the prediction from the variance equation. Given that the homoscedasticity assumption is not met, there is further evidence of preferring the logistic model to the linear regression.

A final drawback of the linear model is that, as shown in its name, it only considers linear combinations of the input variables to predict the output variables. However, the logistic regression model also follows this linear combination form, and so is not any more or less desirable for this reason alone. However, given that in this scenario the previous two drawbacks were so large, the logistic regression model should be preferred instead.

Part 5 (10 points)

Interpret the main result of your final model in terms of both odds and probability of failure. Summarize the final result with respect to the question(s) being asked and key takeaways from the analysis.

The final model is the logistic regression which uses temperature as the single explanatory variable. The model is of the form

$$\text{logit}(\hat{\pi}) = \hat{\beta}_0 + \hat{\beta}_1(\text{Temp})$$

22

```
beta_0_hat <- fail_model$coefficients[1]
beta_1_hat <- fail_model$coefficients[2]
```

The intercept in the model, β_0, had the value 15.0429016 while the coefficient of `Temp` had the value -0.2321627. The coefficient on the `Temp` variable can be interpreted as decreasing the odds of any O-ring failures by 0.7928171 with every 1°F increase in temperature and all else constant.

This is a key finding. It underscores the importance git status of atmospheric temperature on the day of the launch, and that lower temperatures would increase the odds of an O-ring failure occurring. This is supported by Dalal et al. when they state the resiliency of an O-ring is highly dependent on temperature.

On the day of the launch the temperature was 31°F. At this point, the predicted the probability of any O-ring failure is 0.9996088, with a 95 % confidence interval (LRT) as 0.8036982 and 1.

Therefore, with the 95% confidence interval so high, at least one O-ring is very likely to fail under these conditions. If this was known, the launch could have been postponed for warmer weather when there was a lower probability of a failure. We can conclude that the catastrophe could have been avoided had NASA managers appropriately analyzed the data by factoring in both the successful and failed launches.

Citation:

[1] Dalal, Siddhartha R., et al. "Risk Analysis of the Space Shuttle: Pre-Challenger Prediction of Failure." Journal of the American Statistical Association, vol. 84, no. 408, [American Statistical Association, Taylor & Francis, Ltd.], 1989, pp. 945–57, https://doi.org/10.2307/2290069.

[2] Presidential Commission on the Space Shuttle *Challenger* Accident (1986), *Report of the Presidential Commission on the Space Shuttle Challenger Accident* (Vols. 1 & 2), Washington, D.C.: Author.