W271 Group Lab 1

Investigating the 1986 Space Shuttle Challenger Accident

Nathan Martinez, Meera Sharma, Hannah George, and Haile Bizunehe

Contents

Abstract

This study re-analyzes data from the 1986 Challenger accident to replicate an analysis performed by Dalal, Fowlkes and Hoadley (1989) and to use various statistical techniques to test other aspects of the flight data. In our analysis, we found that only the variable temperature had a statistically significant effect on the number of O-ring failures. We further look into the Challenger crash and whether the temperature was likely to be the cause of it. Using various methods, we determined that low temperature results in a higher probability of O-ring failure. However, given the small amount of data we had on low temperatures and O-ring failures, we should be skeptical of the results.

1 Introduction

1.1 Research question

Our research question is:

How does the expected number of O-ring failures change with varying temperature and pressure?

2 Data (20 points)

```
summary(df)
```

```
##      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
```

The summary function gives us a quick overview of the data. From this we can tell that `Flight` appears to be an ID column, taking on values 1 through 23 to identify the particular row of the data. `Number` tells us the number of O-rings on the shuttle, this value is always 6. `O.ring` is the variable that describes the number of O-ring failures for the particular flight. This variable is typically 0 - most flights do not have an O-ring failure - however, for those that do, there tends to only be 1 to 2 failures out of the possible 6 total. The `Temp` variable tells us the temperature at the time of the flight, with a minimum value of 53 and a maximum value of 81. `Pressure` variable tells us the pressure at the time of the flight, with a minimum value of 50 and a maximum value of 200.

2.1 Description

2.1.1 Part 1

The data in this analysis is used from 23 out of 24 space shuttle orbital flights between April 12, 1981, and January 12, 1986, which covers all flights except one before the Space Shuttle Challenger disaster on January 28, 1986. After each launch, the rocket motors were recovered from the ocean for inspection and potential reuse. One flight data is missing as the booster sets were lost in the ocean and it was not possible to retrieve the data and inspect it.

The data consists of temperature for the joints connecting the Solid Rocket Motor casings, propellant gas pressure, number of distressed thermal O-rings used to seal the joints, and the total number of O-rings. Our dependent variable is the number of distressed thermal O-rings used to seal the joints.

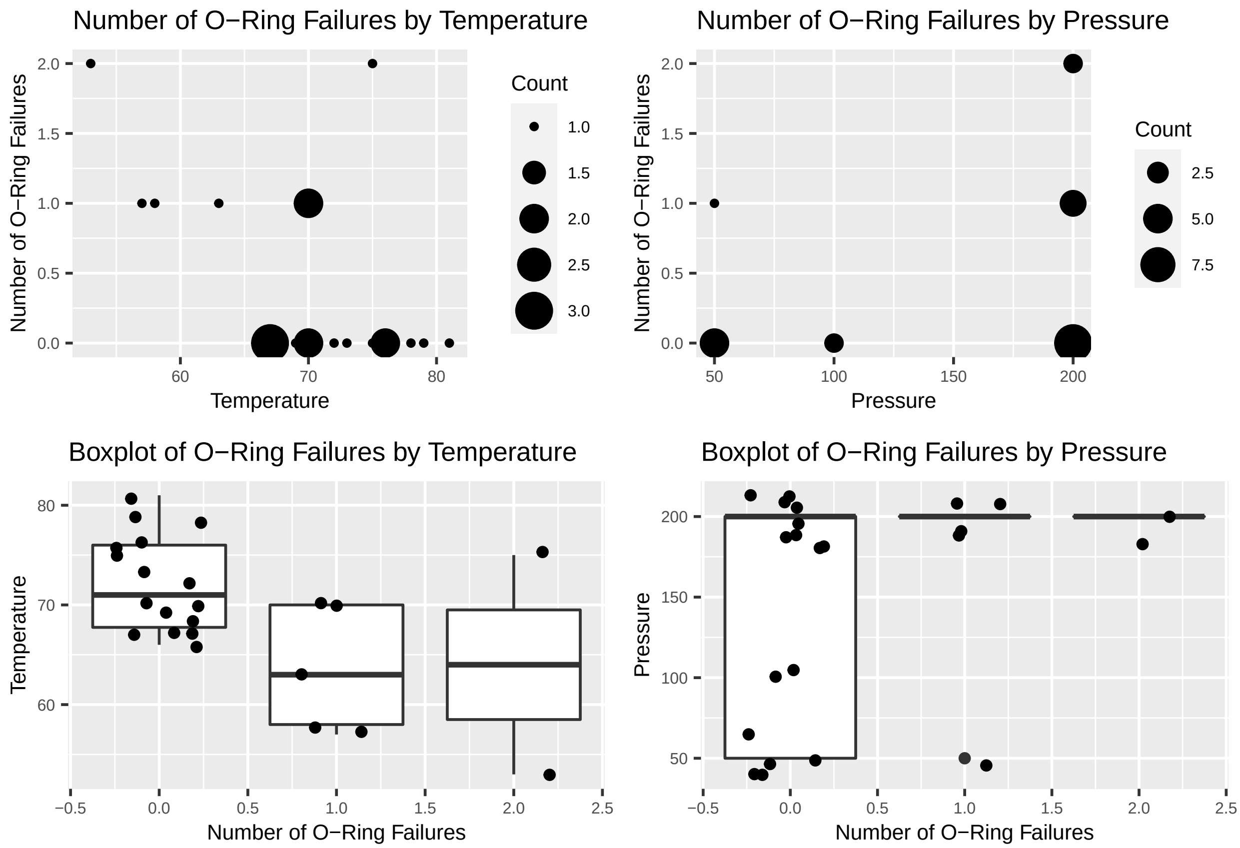
Our population is all O-rings used for space shuttle orbital flights before January 28, 1986 that use the improved version of Solid Rocket Motor Casing Joint Design that accommodates larger-sized shuttle rocket motors and has a second O-ring. The whole population is used in this analysis report. It is important to note that joint temperatures were between 53°F and 81°F for the 23 space shuttle orbital flights, while the temperature on January 28, 1986 was 31°F.

2.1.2 Part 2

By taking independent assumptions for all flights, we can treat the number of distressed thermal O-rings as a binomial random variable and satisfy one of the requirements of a binomial regression that we will use for modeling.

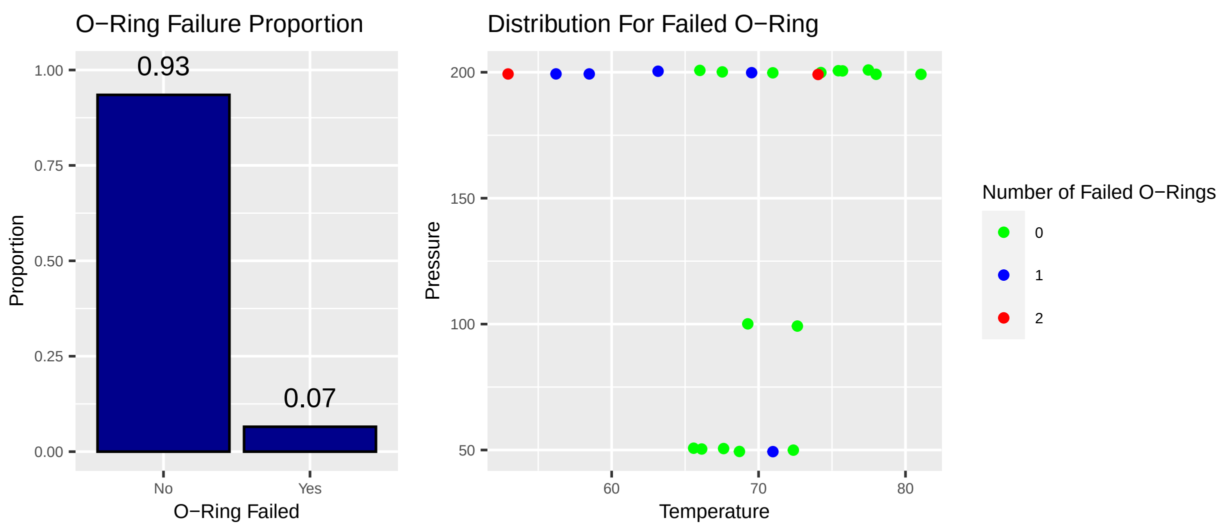
The potential problem with the independence assumption is that if any O-ring is experiencing thermal distress, there is a high chance that other O-rings in the same space shuttle have similar distress. Second, there may be design changes on any of the four-vehicle fleets or how the flight is performed based on previous flight performance. If any of these changes happen, a relation between flights exists, and the previous flight affects the subsequent flight. An existence of dependency between the outcome can lead to substantial bias in estimated standard errors.

2.2 Key Features



Based on the scatter plots above, we can see there is a negative relationship between temperature and the number of O-ring failures, that is, as temperature increases the number of O-ring failures appears to decrease. This intuition is confirmed when we look at the box plot, with both the mean and quantiles for temperature decreasing as number of O-ring failures increases.

However, there does not appear to be much of a relationship at all when we look at pressure. The scatter plot appears randomly distributed and the box plot is non descriptive at best.



The above figure shows that in general the failure rate of O-ring is very low. In the joint distribution of temperature and pressure marked by the number of thermal distress on the O-Ring. Most of the distress happened at a pressure of 200 and temperatures 63 and below degrees Fahrenheit. The pressure does not give any additional clue on if it has a role in thermal distress as the joint distribution did not expand what we already learned from the previous figures. Note that jitter is used on the scatter plot to show multiple points layered together.

3 Analysis

3.1 Reproducing Previous Analysis (10 points)

3.1.1 Part 1

To build a binomial logistic regression model, we used aggregated data where O.ring / Number gives us an observed proportion of O-ring failures for each orbital flight and the weights argument to specify the number of O-rings per orbital flight.

```
model_1 <- glm(formula = O.ring / Number ~ Temp + Pressure, data = df,
               family = 'binomial', weights = Number)

Anova(model_1, test = 'LR')
```

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

Based on the results of the likelihood ratio test, and the EDA performed above, the pressure variable does not appear to have either a statistically significant or practically significant effect on the number of O-ring failures. However, temperature does appear to be significant and should be used in whichever model we create.

3.1.2 Part 2

Based on the results of the likelihood ratio test above, we would also conclude that the pressure variable should not be included in the model. The effect of the variable is not statistically significant, and the scatter plot of pressure versus number of O-ring failures shows no correlation between the two.

3.2 Confidence Intervals (20 points)

3.2.1 Part 1

```r
# 1. Estimate the logistic regression model.
model_2 <- glm(formula = O.ring / Number ~ Temp, data = df, family = 'binomial',
               weights = Number)
summary(model_2)
```

```
##
## Call:
## glm(formula = O.ring/Number ~ Temp, family = "binomial", data = df,
##     weights = Number)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.95227  -0.78299  -0.54117  -0.04379   2.65152
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  5.08498    3.05247   1.666   0.0957 .
## Temp        -0.11560    0.04702  -2.458   0.0140 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 24.230  on 22  degrees of freedom
## Residual deviance: 18.086  on 21  degrees of freedom
## AIC: 35.647
##
## Number of Fisher Scoring iterations: 5
```

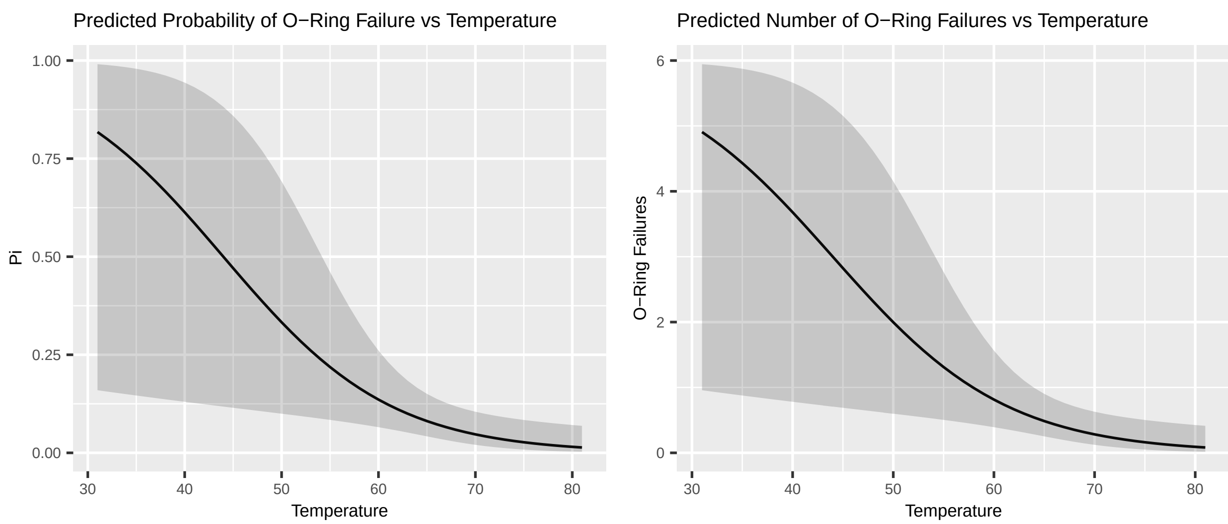
```r
# 2. Determine if a quadratic term is needed in the model for the temperature in
#    this model.
model_3 <- glm(formula = O.ring / Number ~ Temp + I(Temp ^ 2), data = df,
```

```
              family = 'binomial', weights = Number)

Anova(model_3)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: O.ring/Number
##            LR Chisq Df Pr(>Chisq)
## Temp        0.71878  1     0.3965
## I(Temp^2)   0.49470  1     0.4818
```

After looking at the likelihood ratio tests for a quadratic term, we would conclude that it is not necessary to include in the model. For the following purposes, we will use model 2.



The confidence intervals get much larger as temperature decreases, this is because the training data did not have too many low-temperature examples. Leading the model to be more uncertain at lower temperatures.

3.2.2 Part 2

When the temperature is 31°, the estimated probability of O-ring failure is 0.8178 (0.1596, 0.9907) and the estimated number of O-ring failures is 4.9066 (0.9576, 5.9439). This confidence interval is fairly wide, but an O-ring failure does appear to be more likely than not. In order to believe this inference we have to have faith that the model is accurately making predictions for data that it doesn't have any examples of. The data the model trained on doesn't have any examples of a temperature this low so it's just speculating based on the trend from higher-temperature data. This speculation assumes a linear trend, which there just isn't a lot of data to back that assumption. So altogether we would be skeptical of the model's predictions.

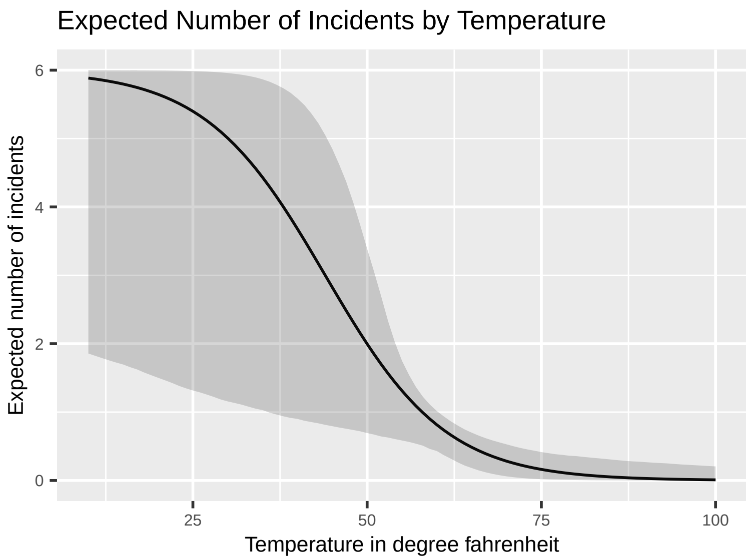
3.3 Bootstrap Confidence Intervals (30 points)

```r
incidents.estimator <- function(b, seed=42, number=6) {
  # Setting seed to make the plot exactly reproduced
  set.seed(seed)
  n <- 23 # Resampled size
  estimation <- array(data = NA, dim=c(91, b))

  # Performing re-sampling repeatedly "b" times
  for(i in 1:b) {
    samples <- sample(x = 1:n, size = n, replace = TRUE)
    model   <- glm(formula = O.ring / Number ~ Temp, data = df[samples,],
                   weights = Number, family = 'binomial')
    # Performing prediction for 10-100° Fahrenheit using each fitted model
    preds <- predict(object = model, newdata = data.frame(Temp = 10:100),
                     type = "response")
    j <- 0
    for(p in preds) {
      j <- j+1 # Index
      estimation[j, i] <- p
    }
  }
  # Fitting a model with the original data for prediction
  model <- glm(formula = O.ring/Number ~ Temp, data = df,
               weights = Number, family = 'binomial')
  # Estimating Expected number of incidents along with 90% CI
  incidents <- data.frame(Temperature = integer(), Incidents = double(),
                          CI.Upper = double(), CI.Lower = double())
  preds <- predict(object = model, newdata = data.frame(Temp = 10:100),
                   type = "response")
  i <- 0
  for(p in preds) {
    i <- i+1 # Index to locate data in the data.frame
    # Getting the 90% CI and multiply by 6 to get estimation for a single flight
    ci <- quantile(estimation[i,], probs = c(0.05, 0.95), names=FALSE)
    incidents[i, 1] <- i + 9 # Temperature
    incidents[i, 2] <- p*number
    incidents[i, 3] <- ci[1]*number
    incidents[i, 4] <- ci[2]*number
  }

  incidents
}

# Performing predictions for 10°-100° temperature and plotting the result
predicted.incidents <- incidents.estimator(5000)
```



We are using the variability among the models fitted using re-sampled data to create a bootstrap distribution for incident prediction and use it to estimate the confidence interval. To build the distribution, we made 23 random draws with replacement (maintaining independence) to create a sample using original observations as a population. We performed model fitting using the sample and prediction at different temperature levels, enabling us to build confidence intervals at each temperature level. The process is replicated 5,000 times to give us a stable distribution that we can use to form a confidence interval. The resulting confidence interval highly varies depending on the temperature. As the temperatures decrease, the confidence interval tends to get wider as it gets further away from the center of the data, causing the uncertainty of the model to become significant.

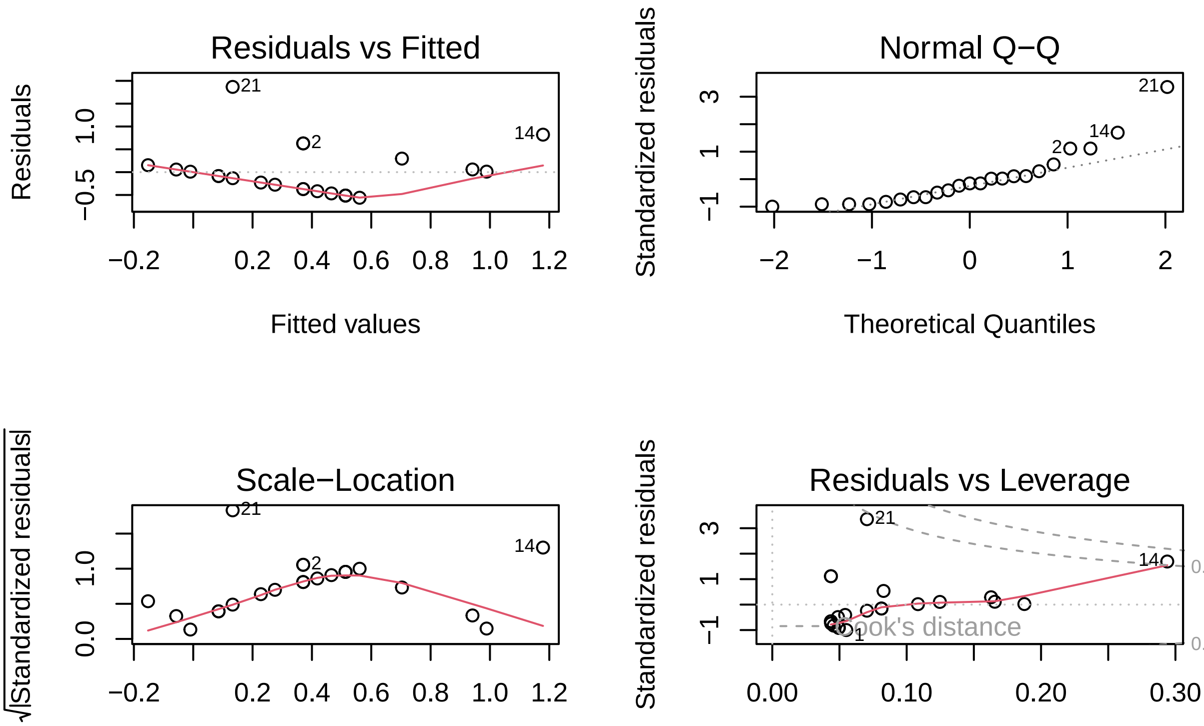
3.4 Alternative Specification (10 points)

```r
mod.linear <- lm(O.ring ~ Temp, data = df)
summary(mod.linear)
```

```
##
## Call:
## lm(formula = O.ring ~ Temp, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5608 -0.3944 -0.0854  0.1056  1.8671
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.69841    1.21951   3.033  0.00633 **
## Temp        -0.04754    0.01744  -2.725  0.01268 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.5774 on 21 degrees of freedom
## Multiple R-squared:  0.2613, Adjusted R-squared:  0.2261
## F-statistic: 7.426 on 1 and 21 DF,  p-value: 0.01268
```

The Temperature is statistically significant and negatively related to the number of O-ring failures for a given flight. For every 1-degree Fahrenheit increase, the number of O-ring failures decreases by 0.04754.



In the residuals vs fitted plot, there is some non-linear relationship that was not explained by the model and left out in the residuals. In the Q-Q plot, the residuals are supposed to follow a straight line to be normally distributed, but it deviates by forming a curve which might be a result of skewed data. In the scale-location plot, which shows us if there is equal variance (homoscedasticity), we see that the residuals are not randomly spread, and because of that, the smooth red line is not horizontal. In the Residual vs Leverage plot, which shows us that if there are influential observations. All the observations are within Cook's distance lines, with one observation 14 on the borderline. We can say that there are no influential cases that can alter the slope coefficient significantly if removed.

Assessments for linear regression model assumptions:
The first assumption is that the data is generated using the Independent and Identically Distributed (IID) sampling process to ensure that the data comes from the same process and that conclusions can be made from a sample to the full population. For this analysis, we used the entire population needed to asses the Challenger disaster on January 28, 1986, and it satisfies the identically distributed requirement. There is a potential for each flight to have a dependency, but there is no evidence that the flight operation and the design is changed based on previous flight performance; as a result, we can assume that they are independent. The second assumption is linearity (Linear Conditional

Expectation), which is one of the three conditions we need to meet under CLM to make our OLS regression model produce unbiased estimates. In order to evaluate the linearity, we examined the model's residuals against the Temp explaining variable, and we can see that the linear relationship does not exist and the assumption is not satisfied. The third assumption is no perfect collinearity where there is no exact linear relationship among the independent variables. Our model has only one variable, and this assumption is satisfied. The fourth assumption is zero conditional mean error which is diagnosed using residuals vs fitted values in the above figure to evaluate if the error (residual) of the model is centered on prediction. In the plot, we see a non-linear relationship that needs to be modeled. As a result, the assumption is not satisfied. The fifth assumption is homoskedastic errors where we have constant error variance across the entire range of the x's. Using occular test, we can see that homoskedastic errors assumption is not satisfied. The sixth assumption is normally distributed errors which is not the case as shown on the Q-Q plot.

The linear regression violated many assumptions, and its prediction can go negative or above 6, which is invalid. Using logistic regression, we do not need to assume a linear relationship between the explanatory variable and response variable, normally distributed residuals and residuals to have constant variance. As a result, binary logistic regression is better than linear regression.

4 Conclusions (10 points)

This study re-analyzes data from the 1986 Challenger accident to replicate an analysis performed by Dalal, Fowlkes and Hoadley (1989) and to use various statistical techniques to test other aspects of the flight data. We developed several logistic regression models to test the relationship between the dependent variable, number of O-ring failures, and the explanatory variables of temperature, pressure and a squared quadratic term for temperature. After performing likelihood ratio tests on the various models, we show that temperature is the only variable that is statistically significant. Specifically, we show that a decrease of 1-degree°F is associated with a 0.11 percentage point increase in O-ring failure. Similarly, we show that the odds of 0-ring failure increase by 12% (1.12) for a 1-degree°F decrease in temperature.

Although the Challenger crash occurred when it was 31°F, the dataset only includes joint temperatures between 53°F and 81°F. Using our preferred model, and the existing dataset, we show that when temperature decreases to 31°, the estimated probability of O-ring failure increases to 82% with a 95% confidence interval between 16% and 99%. When using a bootstrapping technique on the 23 flight samples, we show that the number of O-ring failures increases with colder temperatures and the confidence interval of the number of O-ring failures range from a little over 1 to 6 when temperatures decrease to 30°F. These results make it apparent that flying at colder temperatures is extremely risky but it also shows with the wide confidence interval that we should be skeptical of model predictions at lower temperatures.