# W271 Group Lab 1

### Investigating the 1986 Space Shuttle Challenger Accident

Michael Denton, Srila Maiti, Olivia Pratt, Emily Robles, Elizabeth Willard

## Contents

**Abstract**

This report will, indeed, be abstract. No, instead, describe your goals your approach, and what you learn.

# 1 Introduction

## 1.1 Research question

## 2  Data (20 points)

### 2.1  Description

The data-set for analyzing the risk of O-ring failure in the Challenger Space Shuttle flight consists of twenty-three (23) out of the twenty-four (24) prior launches (one launch data is not available because the motors were lost at sea). The flight number increases by increment of one and corresponds to the incremental number of the flight (increases over date-time). The data is generated by the recovery of the solid rocket motors and subsequent inspection of each component.

We assume that each flight is independent and identically distributed (IID) for the purposes of this analysis. We are able to make this assumption because each flight was, theoretically, in the same configuration with O-rings manufactured, spec'd and installed in the same way. Without the assumption of independence, our ability to infer meaning from the variables would be suspect and we would not be able to conclude with certainty from the statistical analysis. However, the reality is that certain tolerances exist for everything, and parts may not be completely the same with each flight. In fact, the original author (Dalal et al) also expressed that the engineering knowledge of part tolerances was not necessarily consistent with the sampling distribution and conducted a thorough analysis to allevieate concerns.

In addition to the Flight number, variables for launch temperature (**Temperature (°F)**), leak test pressure (**Pressure (psi)**), and the number of O rings having some thermal distress (**O-Ring**) are recorded. The first and last values data set is presented in **Table 1**.

Table 1: First Five Data Points of the Measured Flights Prior to Challenger Disaster

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

### 2.2  Key Features

Evident in in the dataset are some peculiarities corresponding to the available features. The first is that, over time, the leak test pressure changes. Dalal et al. explain that this was due to finding that the putty by itself could sustain 50 psi, so the leak test was increased to 100, and then to 200 psi to properly test for O ring capability. It's also noted, however, that this increased leak pressure may have contributed to blow holes, which can lead to erosion in the putty.

Another major finding is the temperatures that were experienced leading up to the fatal Challenger flight, which was scheduled to launch with a temperature hovering around freezing (31 °F). As seen in **Table 2** explicitly, the minimum temperature experienced prior to the launch failure was 53°F. It's also seen in **Table 1** that this particular launch corresponded to secondary O ring distress and and the mechanism was reportedly primary O ring blow by. The presence of secondary O-ring

failure is critical because it shows that this condition could lead to failure in both O rings and thus penetration of the engine gases.

Table 2: Baseline Statistics of Prior Flights to Challenger Disaster

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| Flight | 23 | 12.000 | 6.782 | 1 | 23 |
| Temp | 23 | 69.565 | 7.057 | 53 | 81 |
| Pressure | 23 | 152.174 | 68.221 | 50 | 200 |
| O.ring | 23 | 0.391 | 0.656 | 0 | 2 |

The temperature effect alluded to in the data set is confirmed in **Figure 1**, which is a panel figure comparing the features and their reported Pearson correlation coefficients. It's shown that the correlation between the the temperature component and the O-ring failures is -0.51, meaning that lower temperatures were considered to result in more O-ring failures. Also note the correlation between *Pressure* and *Flight*, which is easily explained by the change in procedure over time. To a lessor degree also exists a correlation between *Pressure* and *O ring* which may indicate a potential pressure effect to explore. This is also supported by the theory discussed above that high leak test pressure can lead to putty erosion. However, Dalal et al largely dismissed *Pressure* as an important explanatory variable.
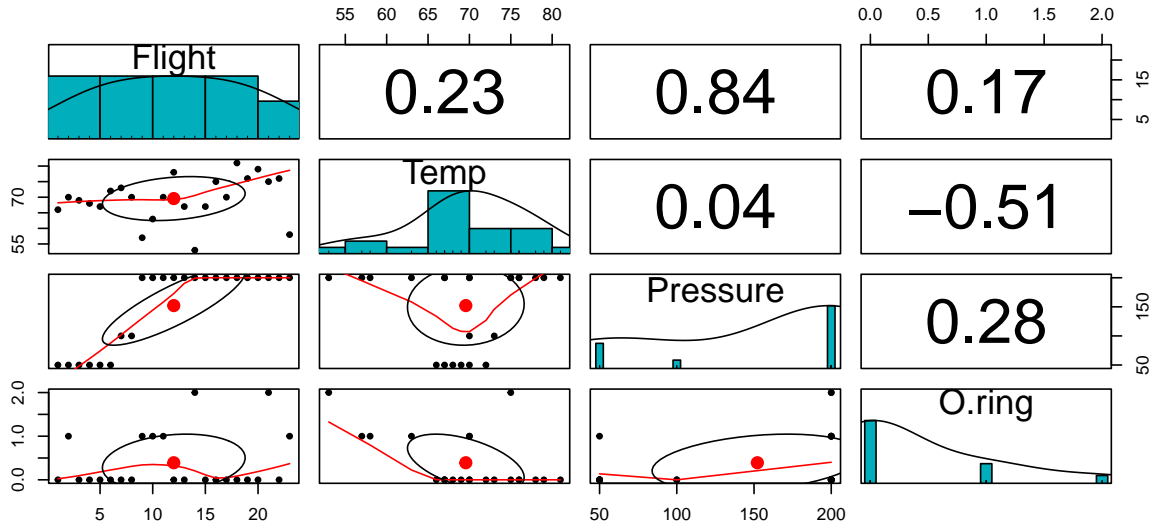


Figure 1: Panel Plot of Flight Variables with Pearson Correlation

# 3 Analysis

## 3.1 Reproducing Previous Analysis

From the exploratory data analysis, we found negative correlation between *Temp* and *O-ring* and positive correlation between *Pressure* and *O-ring.*We want to include both explanatory features in our first logistic regression model. For a given launch $i$, we denote the probability for an O-ring to fail as $\pi_i$, launch temperature as $t_i$ and leak test pressure as $p_i$. With that, the binomial logistic regression model will be formulated as:-

$$logit\left(\pi_i\right) = log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 t_i + \beta_2 p_i$$

```
# Binomial Logit Model
model_binomial_lr <- glm(formula = O.ring / Number ~ Temp + Pressure,
                data = df, family = binomial(link = logit), weights = Number)
# Confidence interval
model_binomial_lr_confint <- confint(model_binomial_lr)
```

For the binomial logistic regression model, we have included two explanatory numeric variables *Temp* and *Pressure*. We have included them in the as-is form without any transformation. The outcome variables is a proportion, derived as the ratio of number o-ring failure to total number of o-ring present, which is 6. The estimated model is shown in **Table 3** along with 95% confidence intervals and presented below:

$$logit\left(\hat{\pi}\right) = 2.520 - 0.098\text{Temp} + 0.008\text{Pressure}$$

We can also conduct a Likelihood Ratio test using the *Anova* function from the *car* package in R to determine statistical importance of each explanatory variable. That test, for *Temperature* yields a $\chi^2$ value of 5.18 corresponding to a P value of 0.02, which is significant. For *Pressure*, the test yields a $\chi^2$ value of 1.54 corresponding to a P value of 0.21, which is not significant.

### 3.1.1 Major Observations:

- In the results of the likelihood ratio test, we see that temperature appears to be statistically significant predictor on the proportion of o-ring failures. It is also showing the inverse relationship between o-ring failure proportion and the temperature.
- The pressure variable does not appear to be statistically significant predictor for predicting of o-ring failure proportion.
- The coefficient of $t_i$ is estimated to be -0.0982968 with the 95% Wald confidence interval of -0.1940717 to -0.0135629, indicating that the decrease on the launch temperature would cause the increase on the odds for an O-ring to fail.
- In Anova test, involving feature variable of *Temp*, according to $H_0 : \beta_1 = 0$ and $H_\alpha : \beta_1 \neq 0$. The LRT statistic for *Temp* is 5.1838035 with a p-value of 0.0227984.
- Using the Type I Error rate $alpha = 0.05$, we reject the null hypothesis and accept that the feature *Temp* is important to be included in the model with other feature variable *Pressure* is included the model.

- For the test of *Pressure* with $H_0 : \beta_2 = 0$ vs. $H_\alpha : \beta_2 \neq 0$, The LRT statistic of *Pressure* is 1.5406572 with a p-value of 0.21452. Using the Type I Error rate $alpha = 0.05$, we fail to reject the null hypothesis. Therefore, there is a lack of evidence to include the feature variable *Pressure* in the model.
- With 10 deg F temperature drop increases the odds o-ring failure by approximately 67%.

2. Dalal, Fowlkes, and Hoadley (1989) chose to remove `pressure` from the model based on their likelihood ratio tests. Critically evaluate, using your test results and understanding of the question and data, whether `pressure` should be included in the model, or instead, `pressure` should not be included in the model. Your report needs to make a determination, argue why it is most appropriate choice, and make note of how (if at all) the model results are affected by the choice of including or excluding `pressure`.

For the test of *Pressure* with $H_0 : \beta_2 = 0$ vs. $H_\alpha : \beta_2 \neq 0$, The LRT statistic of *Pressure* is 1.5406572 with a p-value of 0.21452. Using the Type I Error rate $alpha = 0.05$, we fail to reject the null hypothesis. Therefore, there is a lack of evidence to include the feature variable *Pressure* in the model.

Thus, `Pressure` was probably removed in response to it lack of importance detected by the likelihood ratio test and because it does not add much systematic variation to the model (i.e. `Pressure` has unique values 50, 100, 200). Our data set is very small and not representative. We do not have enough examples for PSI 50 and 100. Potential problems could arise by excluding *Pressure* from the model. If more data was available, pressure along with the interaction with temperature might be helpful for explainability. If pressure is correlated with temperature, and o-ring failure, by removing pressure, we would have introduced omitted variable bias.

## 3.2 Confidence Intervals

In our next model, we are just including temperature as our explanatory variable. For a given launch $i$, we denote the probability for an O-ring to fail as $\pi_i$, launch temperature as $t_i$. With that, the model will be formulated as:-

$$logit\,(\pi_i) = \beta_0 + \beta_1 t_i$$

```
# Binomial Logit Model with just temperature
model_binomial_lr_temp <- glm(formula = O.ring / Number ~ Temp,
              data = df, family = binomial(link = logit), weights = Number)
# Confidence interval
model_binomial_lr_temp_confint <- confint(model_binomial_lr_temp)
```

Binomial logistic regression model with temperature only is estimated in **Table 3** and shown here:

$$logit\,(\hat{\pi}) = 5.085 - -0.116 \text{Temp}$$

Again, we perform a likelihood ratio test using *Anova* and find $\chi^2$ value of 6.14 corresponding to a P value of 0.01, which is significant.

Finally, in our third logistic regression model, we use temperature and it's quadratic term as explanatory features to predict the o-ring failure probability. For a given launch $i$, we denote the

probability for an O-ring to fail as $\pi_i$, launch temperature as $t_i$. With that, the model will be formulated as:-

$$logit(\pi_i) = \beta_0 + \beta_1 t_i + \beta_2 t_i^2$$

```r
# Binomial logistic regression with quadratic temperature term
model_binomial_lr_temp_poly <- glm(formula = O.ring / Number ~ Temp + I(Temp ^ 2),
            data = df, family = binomial(link = logit), weights = Number)

# Confidence interval
model_binomial_lr_temp_poly_confint <- confint(model_binomial_lr_temp_poly)
```

A likelihood ratio test is again ran on the polynomial model, with $\chi^2$ value of 0.72 for *Temperature* and P value of 0.4. For the quadratic term, there is a $\chi^2$ value of 0.49 for *Temperature* and P value of 0.48.

A comparison of all three models is presented in **Table 3**.

Table 3: Comparison of Statistical Models to Predict Number of Failures

|  | *Dependent variable:* | | |
|---|---|---|---|
|  | | O.ring/Number | |
|  | Temp + Press | Temp Only | Temp w/ Quadratic |
|  | (1) | (2) | (3) |
| Temp | −0.098** | −0.116** | −0.651 |
|  | (−0.194, −0.014) | (−0.212, −0.024) | (−2.048, 0.952) |
|  |  |  |  |
| Pressure | 0.008 | | |
|  | (−0.004, 0.029) | | |
|  |  |  |  |
| I(Temp^2) | | | 0.004 |
|  | | | (−0.008, 0.015) |
|  |  |  |  |
| Constant | 2.520 | 5.085* | 22.126 |
|  | (−4.323, 9.773) | (−1.010, 11.185) | (−28.876, 67.282) |
|  |  |  |  |
| Observations | 23 | 23 | 23 |
| Log Likelihood | −15.053 | −15.823 | −15.576 |
| Akaike Inf. Crit. | 36.106 | 35.647 | 37.152 |
| *Note:* | | | *p<0.1; **p<0.05; ***p<0.01 |

We wanted to see if the quadratic temperature has any effect on predicting o-ring failure and we do not see any statistical significance of it. So we are not going to use this quadratic temperature feature in our model.

### 3.2.1 Probability Figures

Using the binomial logistic regression model, we predicted the probability of failure using the second model (just *Temp*), and calculated the confidence intervals. Then, we calculated the expected number of O ring failures and plotted that vs. the temperature. This is shown in **Figure 2**
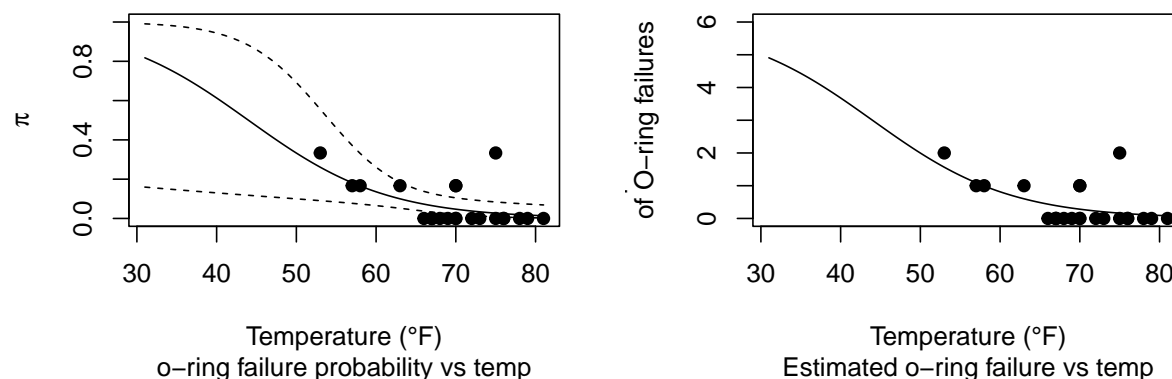


Figure 2: Predicted Probability and Expected Number of O Ring Failures

The confidence band is wider at temperatures lower than 53 °F because there are no observations below 53 °F.

### 3.2.2 Probability of failure at 31°F

Using the models generated above, we predict the probability of a failure occurring on the day of the Challenger disaster, when launch temperature was estimated at 31°F.

```r
# Data to extrapolate
predict.data <- data.frame(Temp = 31)
# Predict surface/link
predict.linear <- predict(object = model_binomial_lr_temp,
                          newdata = predict.data, type = "link")
# Predict response
predict.pi <- predict(object = model_binomial_lr_temp, newdata = predict.data,
                     type = "response")
pfail_31 <- data.frame(estimate=predict.pi, lower=CI_lower_pi[[1]],
                      upper=CI_higher_pi[[1]])
```

The probability of O-ring failure at 31°F is 0.818 with a very wide confidence interval of 0.16, 0.991. These are shown in **Table 4**.

Table 4: Estimated probability of failure with confidence bounds of fatal Challenger Flight.

|   | estimate | lower | upper |
|---|----------|-------|-------|
| 1 | 0.818 | 0.160 | 0.991 |

In order to infer the probability of O-ring failure at 31°F, we need to make the assumption that there is a linear relationship between the log odds of failure of an O-ring and Temperature. Because we use the binomial model, we do not meet the independence assumption. We also need to assume the model is built off a representative data set including records with both high and low temperatures, high and low pressure levels and similar data range is used for both model training and inference. Otherwise the results will be uncertain and we can not trust the model's predictions.

### 3.3 Bootstrap Confidence Intervals (30 points)

For the parametric bootstrap procedure, we have performed the below steps:

- Setting a seed to make the results reproducible.
- Estimate the proportion of o-ring failure using the estimated coefficients.
- Instantiate a result data frame to store the results at 31 deg and 72 deg F.
- Iterate of number of iterations and follow the steps:
  - The original dataset is resampled with replacement to create a new dataset $d$ of size 23.
  - A vector of size 23 with outcome variable $O.ring2$ is generated using rbinom function and using the sampled data and estimated $\hat{\pi}$ values.
  - A new binomial logistic regression model is fitted with the resampled dataset $d$ and outcomes $O.ring2$.

  - The predictions for 31 and 72 °F are found and the estimated probabilities are saved to *results.*

- Finally, 5% and 95% quantiles are derived from the predictions for 31 and 72 °F and stored in *results* and are reported as the 90% confidence intervals.

```
# Set a seed
set.seed(123)
# Use pi estimated from the model
z = model_binomial_lr_temp$coefficients["(Intercept)"] +
    model_binomial_lr_temp$coefficients["Temp"] * df$Temp
pi = exp(z)/(1+exp(z))
# Save the pi array to the dataframe
df$O.ring.pi <- pi
# Dataframe to populate with results
results <- data.frame(pred.31 = numeric(), pred.72 = numeric())
iterations = 10000
for (s in 1:iterations){
  I.sample <- sample(x = 1:nrow(df), size = 23, replace = T)
```

```
# Populate d with samples
d <- df[I.sample,]
# Simulate outcomes using rbinom
# n : sample size, size : number of trials and prob is the probability
O.ring2 <- rbinom(n=23, size=6, prob=d$O.ring.pi)
# Stacking the simulated o.ring value with the sample data
d <- data.frame(d, O.ring2)
# Estimate model with rbinom bootstrap outcomes using the sampled data and sampled data
mod <- glm(O.ring2/Number ~ Temp, family = binomial(link = logit),
           weights = Number, data = d)
# Estimate confidence interval for temp = 31
temp.31.data <- data.frame(Temp=31)
temp.31 <- predict(object = mod, newdata = temp.31.data, type = "response")
# Estimate confidence interval for temp = 72
temp.72.data <- data.frame(Temp=72)
temp.72 <- predict(object = mod, newdata = temp.72.data, type = "response")
results <- results %>% add_row(pred.31 = temp.31, pred.72 = temp.72)
}
```

Table 5: Parametric Bootstrap Analysis Results

|   | temp | lower.CI | upper.CI |
|---|------|----------|----------|
| 1 | 31   | 0.108    | 0.993    |
| 2 | 72   | 0.010    | 0.069    |

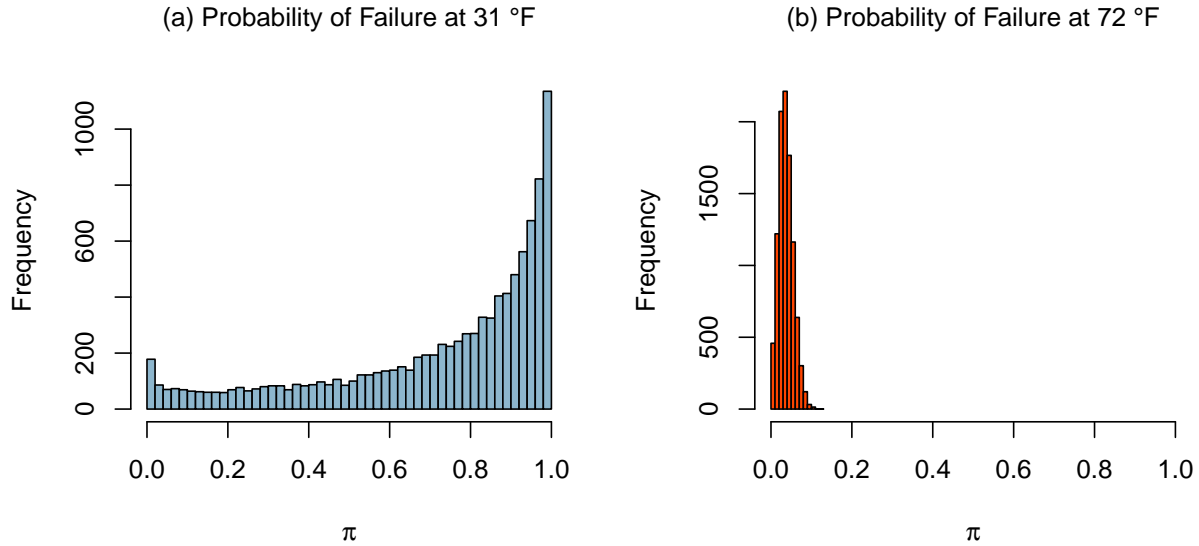(a) Probability of Failure at 31 °F      (b) Probability of Failure at 72 °F



Figure 3: Predicted Probability and Expected Number of O Ring Failures

The parametric bootstrap method to estimate confidence interval also shows that the confidence interval for 31°F is quite wide at 0.1078 and 0.9935, due to the lack of data points at low temperatures. The confidence interval estimated for 72°F is much tighter, at 0.0104 and 0.0694.

Using $10^4$ iterations, through parametric bootstrap process, we have generated the histogram of their probability of failures. The confidence intervals were taken as the 5th and 95th percentile of the distributions. The probability of failures for 31°F took on the full range of 0 and 1, and had a left skew towards 1. In contrast, the probability of failures for 72°F were more concentrated at the left end towards 0.

## 3.4    Alternative Specification

A linear model was built to serve the same purpose of the logistic regression models and predict the number of O ring failures.

```r
# Build Linear Model
model.linear <- lm(O.ring/Number ~ Temp, data = df)
```

Table 6: Linear Model to Predict Number of Failures

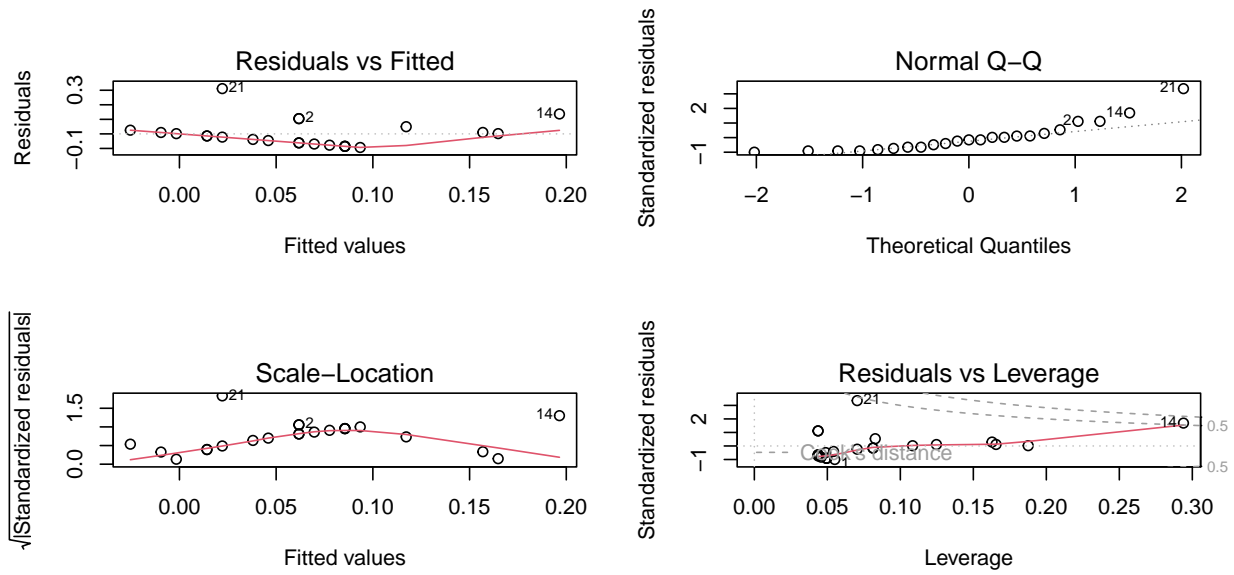|  | *Dependent variable:* |
|---|---|
|  | O.ring/Number |
| Temp | −0.008** |
|  | (0.003) |
| Constant | 0.616*** |
|  | (0.203) |
| Observations | 23 |
| $R^2$ | 0.261 |
| Adjusted $R^2$ | 0.226 |
| Residual Std. Error | 0.096 (df = 21) |
| F Statistic | 7.426** (df = 1; 21) |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

Figure 4: Linear Regression Model

**Histogram of model.linear$residuals**



Figure 5: Linear Regression Model Residual Plot

Estimated linear model is given below:-

$$\frac{O.ring}{Number} = .616 - 0.008 Temp$$

**Observations for the linear model:-** - O-ring failure is inversely related with temperature, that is, with temperature increases, the risk of O-ring failure decreases. - For each unit increase

in temperature, the expected proportion of O-rings that fail drops by -0.0079233 with a standard error of 0.0029075, and *p*-value of 0.0126818.

*Linear regression model assumptions:*

- *IID* For our analysis, we used Challenger dataset and we used the entire population to asses the Challenger disaster on January 28, 1986, and it satisfies the identically distributed requirement. We do not have any evidence that one particular flight record is related to other flight record, thus independent.

- *No perfect colinearity* Our model contains only one explanatory variable, so we satisfy the condition.

- *Linear Conditional Expectations* Based on the given data set, and plot for fitted vs residual, we do not see a linear relationship. This assumption is not satisfied.

- *Homoscadasticity* Homoscadasticity assumption is to have constant residual variance across the range of explanatory variables. The occular test shows that this assumption is not satisfied.

- *Normally Distributed Errors* The relationship between explanatory and the mean of outcome variable is linear. Based on the qqplot, it's obvious that the errors are not normally distributed: we observe that the tail is significantly thinner than a normal distribution. Both Jarque-Bera test value $2 \times 10^{-5}$ and Shapiro-Wilk test value $5.8462539 \times 10^{-61}$ denotes that the distribution on the residuals distribution in question is significantly different from a normal distribution because of the prsence of outliers as seen the residual vs fitted plot.

This linear model formulation also expects the proportion of o-ring failure is linearly related to the explanatory variables for all of their possible values, which is not a valid assumption either.

We would choose binomial logistic regression model over the linear regression model for the below reasons:-

- The binomial logistic regression translates the problem statement in a clear proportion question. The response is the proportion of O-ring failure (e.g. a proportion from 0 to 1).
- The linear model assumptions do not hold true here. Therefore, the linear regression model is not applicable for this problem. On the other hand, using the 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, binomial logistic regression will be a better choice than linear regression.

# 4    Conclusions (10 points)

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