

W271 Lab 1: Challenger O-ring Analysis

Christian Millsop & Tennison Yu

February 3, 2019

Abstract

In this report, a statistical model of the probability of O-ring failure on the space shuttle Challenger is presented. An 81.8% probability of O-ring damage was determined based upon a binomial logistic regression model. Alternative logistic and linear models were also explored. This analysis was inspired by the work of Dalal et al (1989) and agrees in results.

Introduction

On January 28, 1986, the Challenger space rocket suffered catastrophic failure due to O-ring breakdowns inside the solid rocket motors of the booster. As a case study, we emulate the analysis done by Dalal et al (1989) by following questions 4 and 5 found in the exercise section of Chapter 2 of *Analysis of Categorical Data* (2015) by Christopher R. Bilder and Thomas M. Loughin.

The goal of this analysis is to develop a model to predict the probability that the Challenger would fail under the conditions experienced, in particular a temperature of 31F. The dataset being used is a list of damaged O-rings by shuttle flight. Since in the event of failure, examination of the O-rings is impossible, we must use O-ring damage as an indicator of likely failure. If the O-rings withstand enough damage, they will fail.

This report will feature an exploratory data analysis section, followed by modeling and discussion. The latter of which will draw heavily from the exercise questions stated above. We will then conclude by summarizing our findings and presenting potential future work and considerations.

Note that the questions themselves can be found in the appendix and we will refer accordingly to them in square brackets ([]) when we incorporate and address a question. For the questions directly from the assignment, we will use the angled brackets (<>).

Exploratory Data Analysis (EDA)

Setup

The below R snippet is to import the data and the packages we will be using. The data (challenger.csv) can be found here: http://www.chrisbilder.com/categorical/programs_and_data.html

```
library(package = car)
library(package = stargazer)
library(package = Hmisc)
library(package = plyr)
challenger = read.table(file="challenger.csv", sep="," , header = TRUE)
```

Analysis

```
str(challenger)
```

```
## '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 ...
```

Looking at the description above, we are working with 23 rows and 5 columns of data. The columns include information such as

- *Flight*: Flight Number
- *Temp*: Temperature (F) at launch
- *Pressure*: Combustion pressure (psi)
- *O.ring*: Number of primary field O-rings with damage
- *Number*: Total number of primary field O-rings (six total, three each for the two booster rockets)

The response variable is *O.ring*, and the explanatory variables are *Temp* and *Pressure*. By dividing *O.ring* by *Number*, we can also get the percent that were damaged during each flight.

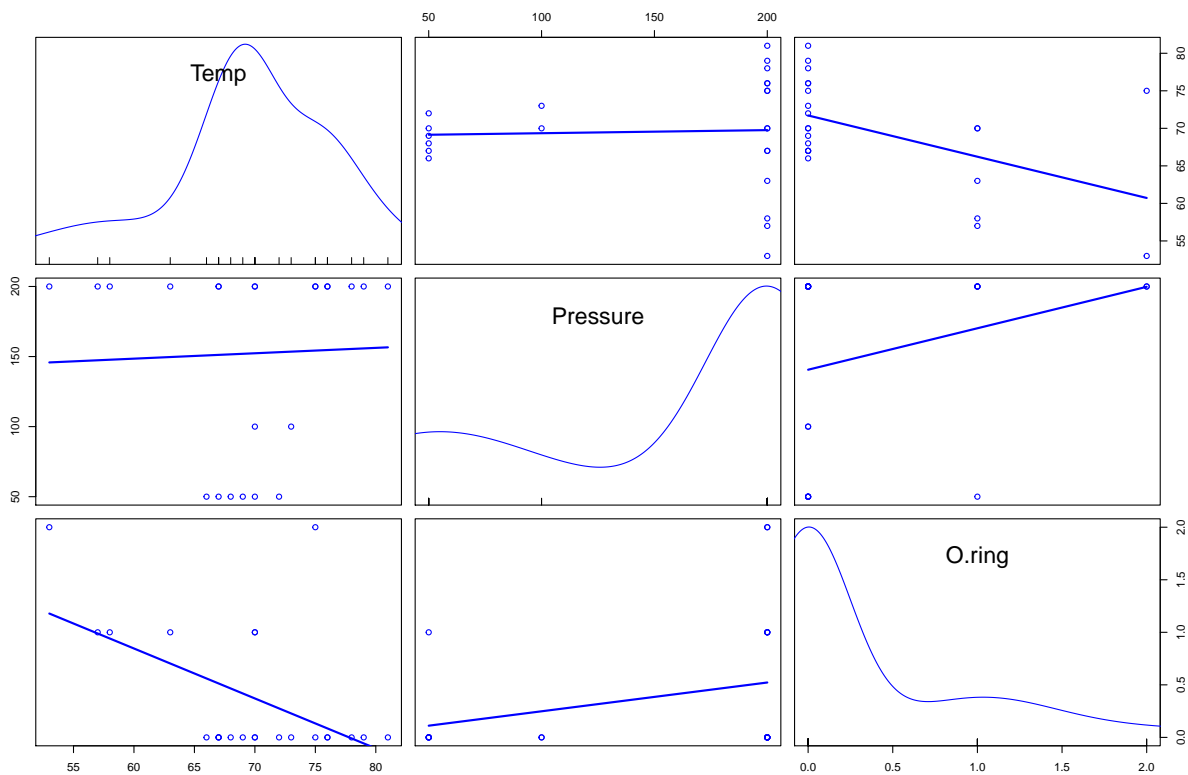
A summary of the explanatory and response variables below shows the general distribution and ranges. We also look at the relationship between the different dependent and independent variables.

```
describe(challenger[,c("Pressure", "Temp", "O.ring")])
```

```
## challenger[, c("Pressure", "Temp", "O.ring")]
##
## 3 Variables      23 Observations
## -----
## Pressure
##      n missing distinct    Info    Mean    Gmd
##      23      0         3  0.706  152.2  67.59
##
## Value      50   100   200
## Frequency    6     2    15
## Proportion 0.261 0.087 0.652
## -----
## Temp
##      n missing distinct    Info    Mean    Gmd    .05    .10
##      23      0        16  0.992  69.57   7.968   57.1   59.0
##      .25   .50    .75    .90    .95
##      67.0   70.0   75.0   77.6   78.9
##
## Value      53   57   58   63   66   67   68   69   70   72
```

```
## Frequency      1      1      1      1      1      3      1      1      4      1
## Proportion 0.043 0.043 0.043 0.043 0.043 0.130 0.043 0.043 0.174 0.043
##
## Value          73      75      76      78      79      81
## Frequency      1      2      2      1      1      1
## Proportion 0.043 0.087 0.087 0.043 0.043 0.043
## -----
## O.ring
##      n missing distinct      Info      Mean      Gmd
##      23      0         3      0.654    0.3913    0.6087
##
## Value          0      1      2
## Frequency      16      5      2
## Proportion 0.696 0.217 0.087
## -----
```

```
scatterplotMatrix(~ Temp + Pressure + O.ring, data=challenger, smooth = FALSE)
```



As above, examination of the data illustrates the general distributions and relationships of the response and explanatory variables.

- *Temp* and *O.ring* have a strong, negative correlation
- *Pressure* and *O.ring* have a weak, positive correlation
- *Temp* observations are clustered around 70.
- *O.ring* damage, in general, is an uncommon event

Background Research

According to Dalal et al (1989), the primary failure mode of the O-rings is by thermal erosion. This occurs when the O-rings are exposed to hot gas that is a product of the rocket reaction.

We also understand that the physical and mechanical properties of the O-rings are highly dependent on the ambient temperature. For instance, mechanically, the compressibility decreases as temperature decreases. Physically, the O-rings also erode very quickly under high temperature. The design was that a putty would be used to protect the O-rings from significant temperature effects of the ignited gas. If the putty were to fail and gases got through via blow holes, the O-rings would degrade and catastrophic failure would ensue.

Pressure also plays a potential role in O-ring damage. Pressure is what causes the putty to seal the O-ring and protect it as mentioned above. In the paper they mention that 50, 100, and 200 psi of pressure were tested but they also mention that extreme pressures can cause the blow holes to form but this isn't very elaborated on. Because there were really only 3 temperatures tested as well, it is rather unclear how significant the effect of pressure is. Nevertheless, under extreme pressure conditions, the O-rings can deform and can be penetrated by the pressurized gas, thus exposing the O-ring to significant thermal erosion as well.

Modeling & Discussion

Logistic Regression Modeling

We seek to develop a model including the terms *Temp* and *Pressure* to predict the probability of O-ring failure. One way to model our data is to use logistic regression with the consideration that the O-rings are thought of as a binomial variable, where each ring is independent of one another and is governed by a probability of success and failure.

However, this is potentially problematic because each of the O-rings are located in a different part of the solid rocket motors (3 sets in each rocket). Variation in the different components and rocket initiation can affect the O-rings on the rocket differently. In terms of the rockets, they can potentially be re-used as well so wear and tear is another factor to consider. Are O-rings replaced each time in such cases? The paper also details the existence of secondary O-rings. These secondary O-rings are meant to be a fail safe in case the primary ones fail. Although it is unlikely since the secondary O-ring doesn't seem to get energized and can lose contact to the tang via joint rotation from the primary O-ring, there could be data in which the secondary O-ring worked and the result of primary O-ring failure was never caught. Lastly, as mentioned above there is also the putty that coats the primary O-ring. The paper mentions its response to pressure but the effects of temperature don't seem well discussed. Under low temperature conditions the O-rings might not seal the cavity within the rocket as designed. Overall, this essentially means that each trial is not likely to be independent of one another and there are likely other variables involved as well. [4a]

To rectify this, the authors instead considered the random variable to be binary/Bernoulli. ie. instead of looking at O-ring failures, they decided to just look at whether the rocket launched successfully or not (0 if no accident/1 if accident). This essentially eliminated the need to for the combinations term in the data and simplified things for them. For the purposes of this exercise though, we stick to the binomial case.

Proposed Models

Based upon the exploratory data analysis and background research, we propose three models:

- A complete model accounting for influence of *Temp* and *Pressure*. [4b]

$$(1) : \text{logit}(\pi) = \beta_0 + \beta_1 \text{Temp} + \beta_2 \text{Pressure}$$

- A reduced model focusing on *Temp* since our background research suggests that this influences performance of the O-rings the most. [5a]

$$(2) : \text{logit}(\pi) = \beta_0 + \beta_1 \text{Temp}$$

- A quadratic model since the distribution of *Temp* vs *O.ring* data suggests that the data could have a non-linear relationship. [5f]

$$(3) : \text{logit}(\pi) = \beta_0 + \beta_1 \text{Temp} + \beta_2 \text{Temp}^2$$

where $\text{logit}(\pi) = \log(\frac{\pi}{1-\pi})$ therefore it is a linking function. π is the probability of failure per O-ring. It is denoted as $p(t, s)$ by Dalal et al (1989) to show the dependency of failure with a given *Temp*, t , and *Pressure*, s . The term $p(t, s)$ also relates to the binomial distribution $P(X = k) = nCk * p(t, s)^k * (1 - p(t, s))^{n-k}$. In the case of no *Pressure* term, s , then $p(t, s)$ is replaced by $p(t)$.

Model Estimates

The code for modeling and the model estimate is below:

```
challenger$prob = challenger$O.ring/challenger$Number

mod.1.full <- glm(
  formula = prob ~ Temp + Pressure,
  family = binomial(link = logit),
  data = challenger
)

mod.1.noPressure <- glm(
  formula = prob ~ Temp,
  family = binomial(link = logit),
  data = challenger
)

mod.temp.squared <- glm(
  formula = prob ~ Temp + I(Temp^2),
  family = binomial(link = logit),
  data = challenger
)
```

```
suppressWarnings(stargazer(mod.1.full, mod.1.noPressure, mod.temp.squared,
keep.stat=c("n", "aic"), omit.table.layout="n", header=FALSE, type='text'))
```

```
##
## =====
##               Dependent variable:
##               -----
##               prob
##               (1)      (2)      (3)
## -----
## Temp          -0.098  -0.116   -0.651
##               (0.110) (0.115) (1.814)
##
## Pressure       0.008
##               (0.019)
##
## I(Temp2)              0.004
##               (0.014)
##
## Constant       2.520    5.085    22.126
##               (8.541) (7.477) (58.284)
##
## -----
## Observations    23      23      23
## Akaike Inf. Crit. 9.229   7.203   9.237
## =====
```

The results show the estimated coefficients, standard errors, and evaluations of the goodness-of-fit, Akaike Inf.Crit. (AIC). The model with the lowest AIC (model 2) should have the best fit vs. parsimony of explanatory variables.

Expressing the estimates in the form of models (1), (2), and (3) above yields:

$$(1) : \text{logit}(\hat{\pi}) = 2.520 + -0.098\text{Temp} + 0.008\text{Pressure}$$

$$(2) : \text{logit}(\hat{\pi}) = 5.085 + -0.116\text{Temp}$$

$$(3) : \text{logit}(\hat{\pi}) = 22.126 + -0.651\text{Temp} + 0.004\text{Temp}^2$$

Interpreting these models, for example model 1, every one unit increase in temperature will decrease the chance of failure by 0.098 units whereas every unit increase in pressure will cause a 0.008 unit increase in chance of failure. This could be understood as low temperature and high pressure affecting O-rings to not seal properly.

Model Selection

Beyond comparing AIC values of the models, we're also interested in the effect that each explanatory variable has on the statistical significance. For example, is model 1 statistically equivalent to model

2. We can do this by comparing the $-2\log(\Lambda)$, where Λ is the likelihood ratio test (LRT), to the $\chi^2_d f$ distribution, allowing us to calculate a p-value.

```
suppressWarnings(anova(mod.1.noPressure, mod.1.full, test = "Chisq"))

## Analysis of Deviance Table
##
## Model 1: prob ~ Temp
## Model 2: prob ~ Temp + Pressure
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      21      3.0144
## 2      20      2.7576  1  0.25678   0.6123

suppressWarnings(anova(mod.1.noPressure, mod.temp.squared, test = "Chisq"))

## Analysis of Deviance Table
##
## Model 1: prob ~ Temp
## Model 2: prob ~ Temp + I(Temp^2)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      21      3.0144
## 2      20      2.9319  1  0.08245   0.774
```

We see that the p-values (0.612 and 0.774) are not significant at $\alpha = 0.05$. This means that all three of our models are statistically equivalent. The residual deviance does have an absolute change between models, but it does not have a statistical effect. Based upon our test results we chose to proceed with the simplest model (2). [4c]

Although parsimony is desirable and we will be proceeding with the simplest model, we do acknowledge that pressure and temperature may have an unexplored interaction effect. Our exploratory data analysis indicated that there isn't much of a relationship between them in the data, but one of the foundation laws of physics is the ideal gas law, $PV = nRT$, which relates pressure (P) and temperature (T) in a gaseous system. Although we are dealing with solid objects, this formula raises the notion that there could be interactive effects. As noted previously high pressure can cause blow holes, which exposes the O-ring to high temperature effects potentially as well and ultimately catastrophic failure. [4d]

Model Predictions

In order to explore the meaning of model 2 and to give it more meaning, we generated plots of π and the expected number of failures along with the 95% confidence interval versus *Temp*. [5b, 5c]

```
alpha = 0.05

ci.plot = function(model, xdata, a) {
  z = qnorm(1-alpha/2)
  predictions = predict(model, newdata=data.frame(Temp=xdata), type="link", se.fit=TRUE)
  lower_ci = predictions$fit - z*predictions$se.fit
  upper_ci = predictions$fit + z*predictions$se.fit
  list(lower=exp(lower_ci)/(1+exp(lower_ci)), upper=exp(upper_ci)/(1+exp(upper_ci)))
}
```

```

}

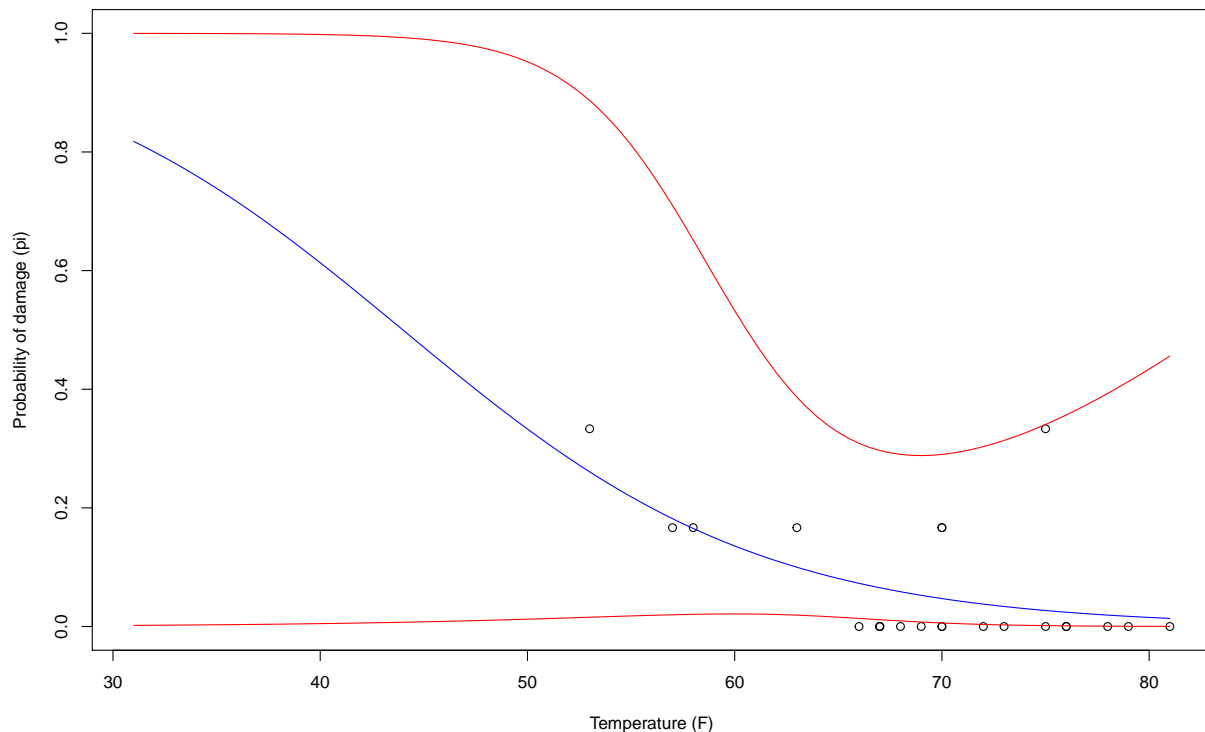
plot(challenger$Temp, challenger$prob, xlim = c (31, 81), ylim=c(0,1),
      xlab="Temperature (F)", ylab="Probability of damage (pi)")

curve (expr = predict (object = mod.1.noPressure,
                       newdata = data.frame(Temp = x), type = "response"),
       col = "blue", add = TRUE, xlim = c (31, 81), ylim=c(0,1))

curve (expr = ci.plot(mod.1.noPressure,x,alpha)$lower,
       col = "red", add = TRUE, xlim = c (31, 81), ylim=c(0,1))

curve (expr = ci.plot(mod.1.noPressure,x,alpha)$upper,
       col = "red", add = TRUE, xlim = c (31, 81), ylim=c(0,1))

```



Looking at the above, it appears that a much safer temperature to launch the shuttle would be around 65F due to the interval gap. Notably, the confidence interval is wider at lower and higher temperatures likely because we did not have as much data at those temperatures to model.

It is evident that this is the case as it can be seen that around 50 degrees, the gap starts to close and it is exactly around this region where we start having data. We can study this a little more by showing a standard error vs *Temp* graph as below for a sequence of temperatures. Notice the dip is in the same region as the confidence interval above. Also the red points indicate the number of values of a specific temperature. As can be seen also, temperatures that are more frequent

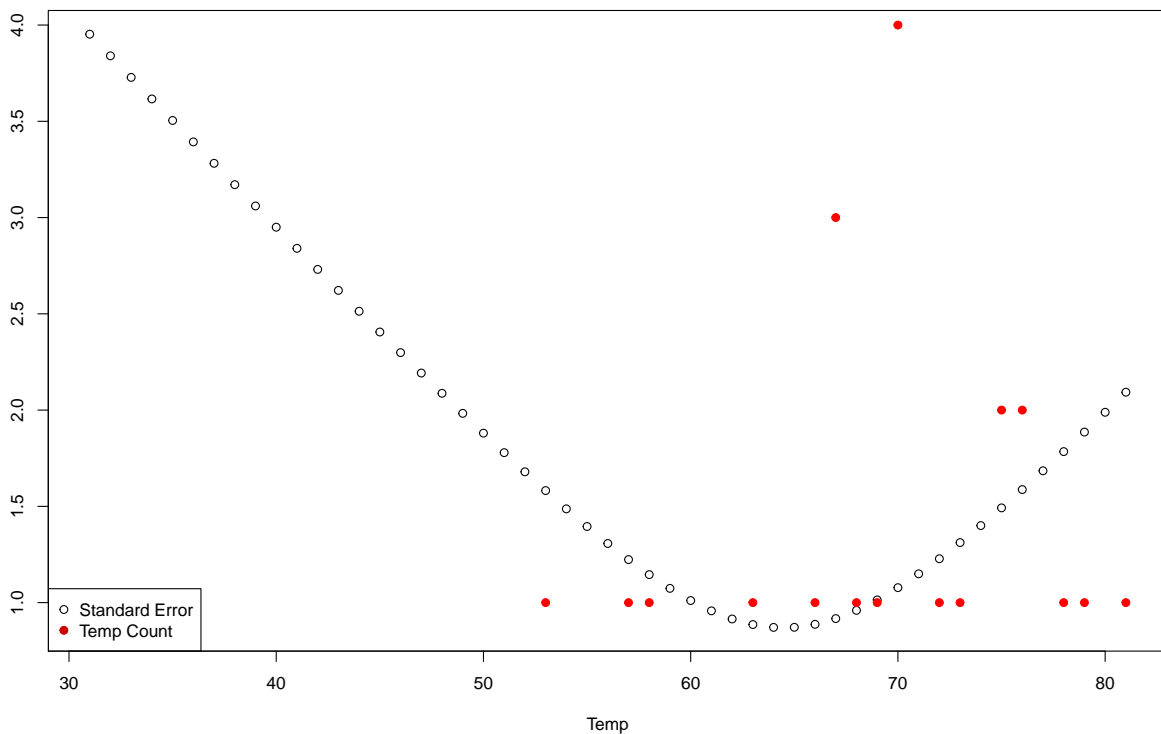
correspond to regions of tighter confidence gaps.

```
temp_range = seq(31,81)
pred_range = predict(object = mod.1.noPressure,
                      newdata = data.frame(Temp=seq(31,81)),
                      type="link", se.fit=TRUE)

plot(x = temp_range,
     y = pred_range$se.fit,
     xlab = 'Temp',
     ylab = ''
)

par(new = T)
with(count(challenger, "Temp"), points(Temp, freq, col = 'red', bg = 'red', pch=19))

legend("bottomleft",
      legend=c('Standard Error', 'Temp Count'),
      lty=c(0,0),
      pch=c(1, 19),
      col=c('black', "red3")
)
```



Under the temperature conditions of Challenger (31F Temperature), our model predicts: **[4d]**

```

predicted_odds = predict(mod.1.noPressure, newdata=data.frame(Temp=31),
                          type="link", se.fit = TRUE)
predicted_prob = exp(predicted_odds$fit)/(1+exp(predicted_odds$fit))
ci_odds = c(predicted_odds$fit - qnorm(1-.05/2)*predicted_odds$se.fit,
            predicted_odds$fit + qnorm(1-.05/2)*predicted_odds$se.fit)
ci = exp(ci_odds)/(1+exp(ci_odds))

cat("Predicted probability of at least one damaged O-ring with 95% confidence interval:\n",
    round(ci[1],4), "<", round(predicted_prob,4), "<", round(ci[2],4))

```

```

## Predicted probability of at least one damaged O-ring with 95% confidence interval:
## 0.0019 < 0.8178 < 0.9999

```

```

cat("\nPredicted count of damaged O-rings with 95% confidence interval:\n",
    round(ci[1]*6,4), "<", round(predicted_prob*6,4), "<", round(ci[2]*6,4))

```

```

##
## Predicted count of damaged O-rings with 95% confidence interval:
## 0.0116 < 4.9066 < 5.9994

```

At 31 degrees, it seems there's about 81.8% chance of damage. However, this result should be taken cautiously. As discussed previously, there is no data below the 50 degree region therefore as validated by the large confidence interval gap, it's hard to say how accurate that zone of prediction is. We would have to assume that below 50 degrees, the behavior is linear per the model and that the effect of pressure does not change or somehow become more dominant in that zone. [5d]

With our model we can also predict the total number of O-rings by temperature that would be damaged. [5b]

```

alpha = 0.05

ci.plot = function(model, xdata, a) {
  z = qnorm(1-alpha/2)
  predictions = predict(model, newdata=data.frame(Temp=xdata), type="link", se.fit=TRUE)
  lower_ci = predictions$fit - z*predictions$se.fit
  upper_ci = predictions$fit + z*predictions$se.fit
  list(lower=exp(lower_ci)/(1+exp(lower_ci))*6, upper=exp(upper_ci)/(1+exp(upper_ci))*6)
}

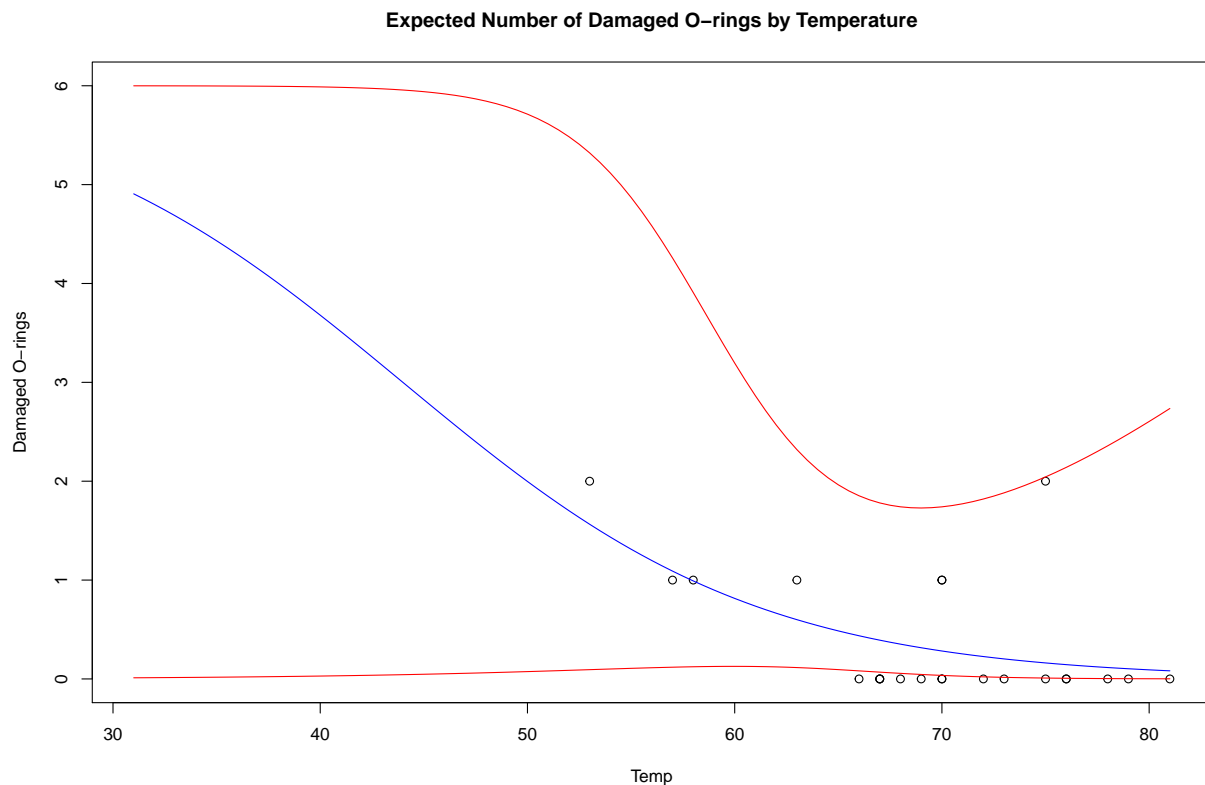
plot(challenger$Temp, challenger$O.ring,
     ylab="Damaged O-rings", xlab = "Temp", ylim=c(0,6), xlim=c(31,81),
     main = "Expected Number of Damaged O-rings by Temperature")

curve (expr = predict (object = mod.1.noPressure,
                      newdata = data.frame(Temp = x), type = "response")*6,
       col = "blue", add = TRUE, xlim = c (31, 81), ylim=c(0,6))

curve (expr = ci.plot(mod.1.noPressure,x,alpha)$lower,
       col = "red", add = TRUE, xlim = c (31, 81), ylim=c(0,6))

```

```
curve (expr = ci.plot(mod.1.noPressure,x,alpha)$upper,
      col = "red", add = TRUE, xlim = c (31, 81), ylim=c(0,6))
```



As can be seen, our model predicts that up to 5 O-rings can fail at low temperatures. This is consistent with the model proposed by the authors.

Bootstrapped Confidence Interval

Given the described weaknesses of the Wald confidence interval, Dalal et al (1989) used an alternative method to calculate confidence intervals. The approach was to bootstrap the models, predict the probability of damage with each model, then select the predictions at the 5% and 95% quantiles to generate a 90% confidence interval. We replicate the bootstrapping method at Dalal's CI of 90% and our model prediction's CI of 95% for comparison. [5e]

```
get_bootstrap_ci = function(df_number, temp, cl){
  vector <- vector("numeric")
  set.seed(1)

  for (i in 1:df_number) {

    df.tmp <- challenger[sample(nrow(challenger), 23, replace = TRUE), ]
    df.tmp$prob = df.tmp$O.ring/df.tmp$Number
```

```

mod.1.tmp <- suppressWarnings(glm(
  formula = prob ~ Temp,
  family = binomial(link = logit),
  data = df.tmp
))

pred.value <- data.frame(Temp=temp)
pred.result <- predict(mod.1.tmp, pred.value, type = "link", se = TRUE)
vector[i] <- exp(pred.result$fit)/(1+ exp(pred.result$fit))
}

results_sorted = sort(vector)

c(Temp=temp,
  CL=paste(c1*100,"%", sep=""),
  CI_lower = results_sorted[df_number*(1-cl)/2],
  CI_upper = results_sorted[df_number*(1-(1-cl)/2)]
)
}

data.frame(rbind(
  get_bootstrap_ci(1000, 31, .9),
  get_bootstrap_ci(1000, 72, .9),
  get_bootstrap_ci(1000, 31, .95),
  get_bootstrap_ci(1000, 72, .95)))

```

```

##   Temp CL          CI_lower          CI_upper
## 1   31 90%  0.164772622913731  0.991365238024074
## 2   72 90%  0.00564553960708654  0.0791598704915577
## 3   31 95%  0.0670588357645897  0.996032089946483
## 4   72 95%  0.00401273535639221  0.0863244236424386

```

The bootstrap results for the 95% CI, $0.0671 < \hat{\pi} < 0.9960$ are slightly narrower than our previous result of $0.0019 < \hat{\pi} < 0.9999$.

The final model is $\text{logit}(\pi) = \beta_0 + \beta_1 \text{Temp}$. In terms of odds, we can see that the odds ratio can be calculated simply as $\frac{\exp(\beta_0 + \beta_1(\text{Temp} + c))}{\exp(\beta_0 + \beta_1(\text{Temp}))}$ which simplifies to $\exp(c\beta_1)$. Thus, because the coefficient of *Temp* is -0.1156, the estimated odds of an O-ring failing decreases by 0.890 times for every c-unit increases in temp. In terms of the probability of failure as seen below, it is very clear that the probability of there being a failure is high when the temperature is low (81.8% at 31 degrees) and this changes as the temperature increases (1.36% at 81 degrees). **<3a>**

Linear Regression Modeling

In the above analysis, we used a combination of logistic regression and binomial theory to calculate a probability and expected number of failures. How does this compare to a traditional linear model? We also perform a classical linear regression to compare. In the linear regression we use the same explanatory variable term, *Temp*, and a slightly different response variable, *O.ring*. In this case,

since the linear model is not bounded from 0 to 1, as in probability, it makes more sense to predict the number of damaged O.rings directly rather than the probability of damage.

```
mod.lin <- lm(

formula = O.ring ~ Temp,
data = challenger

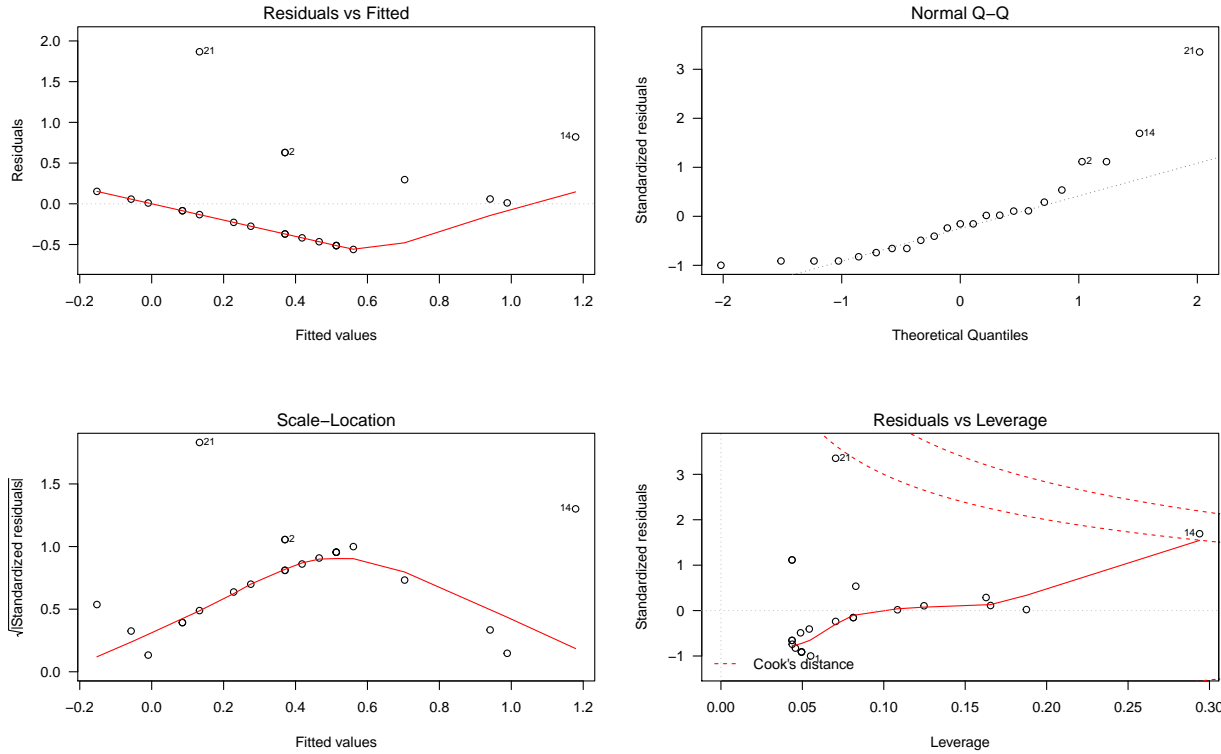
)

summary(mod.lin)

##
## Call:
## lm(formula = O.ring ~ Temp, data = challenger)
##
## 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

par(mfrow = c(2,2), oma = c(0, 0, 1.1, 0))
plot(mod.lin, las = 1)
```

lm(O.ring ~ Temp)



We do not expect the linear model to have good predictive capability. The fit is poor, with an R^2 of 0.2613. Digging further into the diagnostic plots and Classical Linear Model (CLM) assumptions, we see further cause for concern. The residual vs fitted and scale-location plot indicate some sort of non-linear relationship in the data and that the variance is not constant across samples (heteroscedastic). Furthermore, there appear to be outliers. Data points, 14 and 21 are marked above as having considerable leverage. In the data, these were cases where 2 O-rings were damaged instead of either 1 or 0. Because the vast majority of data points are those where only a single O-ring was damaged or none at all, these two data points are skewing the model. In addition like logistic regression, this model would need independence from all descriptive variables and that is not the case as seen with pressure having some kind of effect.

Overall, we prefer to use the binomial logistic model since it is a more conservative model in its approach. A logistic model is advantageous since it does not require a linear relationship between the dependent and independent variables, the residuals do not need to be normally distributed and lastly, homoscedasticity is not required. Ignoring the requirements of either model, we also believe that the logistic model produces a more useful result, ie. probability of O-ring damage. In the context that any O-ring damage can lead to catastrophic failure the probability is more actionable. In other words, its easier to say “what is the probability of damage with these given conditions” via the logistic model versus “What conditions will lead to a prediction of 1 or more”. **<3b>**

Conclusion

We explored three logistic regression models and one linear regression model in this analysis. Based upon statistical examination of these models, we have concluded that the model $\text{logit}(\pi) = \beta_0 + \beta_1 * \text{Temp}$ is most suitable. Our analysis and results are consistent with the finding of Dalal et al (1989).

For logistic model selection we examined the AIC values of different models and the significance of the explanatory variables using LRT. The other models did not explain a greater amount of the variance in the data and were more complicated. This is demonstrated by the AIC value for the chosen model (7.2026) versus the *Pressure* and *Temp*² models (9.2286 and 9.2373). The LRT analysis didn't produce p-values < 0.1 for any variables, but generally showed that *Temp* by itself is the most likely to be significant.

The linear regression model was deemed unsuitable since the fit was poor and the underlying CLM assumptions were not met. The failure of the CLM assumptions reduces our confidence that model would be able to predict O-ring damage in a region of the regression for which there was no data.

References

- [1] Dalal, Siddhartha R., Edward B. Fowlkes, and Bruce Hoadley. "Risk analysis of the space shuttle: Pre-Challenger prediction of failure." *Journal of the American Statistical Association* 84.408 (1989): 945-957.
- [2] Bilder, Christopher R., Loughin, Thomas M.. *Analysis of Categorical Data with R* (Chapman & Hall/CRC Texts in Statistical Science). CRC Press (2015).
- [3] Kelly, Dana L., and Curtis L. Smith. "Risk analysis of the space shuttle: pre-challenger bayesian prediction of failure." *NASA Space Systems Engineering and Risk Management* (2008).
- [4] Simonoff, J. "The Flight of the Space Shuttle Challenger." (2017) <http://people.stern.nyu.edu/jsimonof/classes/2301/pdf/challlog.pdf>

Appendix

Questions from reference [2]

- (4a) The authors use logistic regression to estimate the probability an O-ring will fail. In order to use this model, the authors needed to assume that each O-ring is independent for each launch. Discuss why this assumption is necessary and the potential problems with it. Note that a subsequent analysis helped to alleviate the authors' concerns about independence.
- (4b) Estimate the logistic regression model using the explanatory variables in a linear form.
- (4c) Perform LRTs to judge the importance of the explanatory variables in the model.
- (4d) The temperature was 31ř 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.

5. Continuing Exercise 4, consider the simplified model $\text{logit}(\pi) = \beta_0 + \beta_1 \text{Temp}$ where π is the probability of an O-ring failure. Complete the following:
 - (a) Estimate the model.
 - (b) Construct two plots: (1) π vs. Temp and (2) Expected number of failures vs. Temp. Use a temperature range of 31 to 81 degrees on the x-axis even though the minimum temperature in the data set was 53 degrees.
 - (c) Include the 95% Wald confidence interval bands for on the plot. Why are the bands much wider for lower temperatures than for higher temperatures?
 - (d) The temperature was 31 degrees 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.
 - (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 $\text{logit}(\hat{\pi}) = \hat{\beta}_0 + \hat{\beta}_1 \text{Temp}$; (2) estimate new models for each data set, say $\text{logit}(\hat{\pi}^*) = \hat{\beta}_0^* + \hat{\beta}_1^* \text{Temp}$; and (3) compute $\hat{\pi}^*$ at a specific temperature of interest. The authors used the 0.05 and 0.95 observed quantiles from the $\hat{\pi}^*$ simulated distribution as their 90% confidence interval limits. Using the parametric bootstrap, compute 90% confidence intervals separately at temperatures of 31 and 72 degrees.
 - (f) Determine if a quadratic term is needed in the model for the temperature.

Questions from assignment

- <3a> Interpret the main result of your final model in terms of both odds and probability of failure.
- <3b> 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.