Lab 1: O-Ring Failure Analysis

George Barsukov, Dili Wang

9/27/2019

Intro

In the following analysis, we attempt to accurately model the probability of O-ring failure on the Space Shuttle Challenger based on data gathered from 23 space shuttle launches, some of which contained incidents of failures. Our analysis mainly focuses on using logistic regression to model the probability of single O-ring failure and the binomial probability distribution to ascertain the expectation of number of O-ring failures based on explanatory variables of Temperature and Pressure. In addition, we perform some comparisons between our final model that predicts single O-ring failure to both an alternative logistic regression model that predicts rocket failure, defined by at least one O-ring failure, and a second alternative linear model that predicts the number of O-ring failures using the same explanatory variable of Temperature alone. Our final model predicts the probability of O-ring failure on the day of the Challenger, where Temperature is 31° to be 0.8178 with a 95% Wald confidence interval of (0.1596,0.9907), estimating the probability of one O-ring failure, and therefore rocket failure, to be highly likely.

EDA

```
data <- data.table(read.csv("challenger.csv", header=TRUE, sep=","))
head(data)
```

Based on a summary of the initial reading of the data, we noticed exactly 23 observations. The data contains 2 independent variables, Temp and Pressure, and an integer response variable, O.ring, ranging from 0 to 2. None of the variables contained missing values for any of the observations. The data also included a Number variable, which is the same value of 6 for all observations. Based on the Dalal et al paper, we know that this Number represents the total number of primary O-rings in a rocket, and is a necessary component to our analysis of the binomial logistic regression model.

For the purposes of exploratory data analysis only, we added 2 additional columns of factor transformed versions of O-ring (f_O.ring) and Pressure (f_pressure), both of which have only 3 values and can each be used to categorically partition the data. Please note that neither transformed variable is used in the subsequent code for modeling and analysis and is only specifically used in our EDA to examine the dependence of the response variable on Pressure.

```
#transform response variable O.ring into factor
data[, f_O.ring := as.factor(data$O.ring)]
data[, f_pressure := as.factor(data$Pressure)]
summary(data)
```

Below, we plot the distribution of both explanatory variables, the distribution of the response variable, and a scatterplot showing the relationship between temperature and pressure. It is important to note that the Temperature distibution appears to be symmetric over a range of 53° to 81° and can be approximated by a normal distribution. Due to the appearance of the distribution

and given the small sample size, we are able to use a t-distribution, fit to the mean and standard deviation of the observed temperature values, to sample temperatures for the bootstrap method of confidence interval calclations later on. On the other hand, the Pressure variable shows that there are only 3 pressure values in the data set, where a Pressure of 200 psi is most frequent, followed by that of 50 psi, and lastly by that of 100 psi. The O.ring variable shows the observed number incidents associated with each value of O-ring count decreases as damaged O-ring count increases. Lastly, the scatterplot between Temperature and Pressure does not show any clear relationships and certainly no obvious linearity. The only thing we can conclude from this plot is that for points where Pressure is equal to 200 psi, the Temperature varies greatly and much less so for other values of Pressure, which could also be the result of much fewer observed data points for other values of Pressure.

```
#distribution of temp
temperature = ggplot(data, aes(x = Temp)) +
  geom_histogram(aes(y = ..density..), binwidth = 5, fill="#0072B2", colour="black") +
  ggtitle("Temperature")+theme(plot.title=element_text(lineheight=1,face="bold",size = 12))
#distribution of pressure
pressure = ggplot(data, aes(x = Pressure)) +
  geom_histogram(aes(y = ..density..), binwidth = 50, fill="#0072B2", colour="black") +
  ggtitle("Pressure")+theme(plot.title=element_text(lineheight=1,face="bold",size = 12))
#distribution of O.ring
orings = ggplot(data, aes(x = O.ring)) +
  geom_bar( fill="#0072B2", colour="black")+ggtitle("O-Ring Count")+
  theme(plot.title = element_text(lineheight=1, face="bold",size = 12))
#temperature by pressure
scatter = ggplot(data, aes(x=Temp, y=Pressure))+geom_point()
plot_grid(temperature, pressure, orings, scatter, scale = 1, label_size = 7)
```
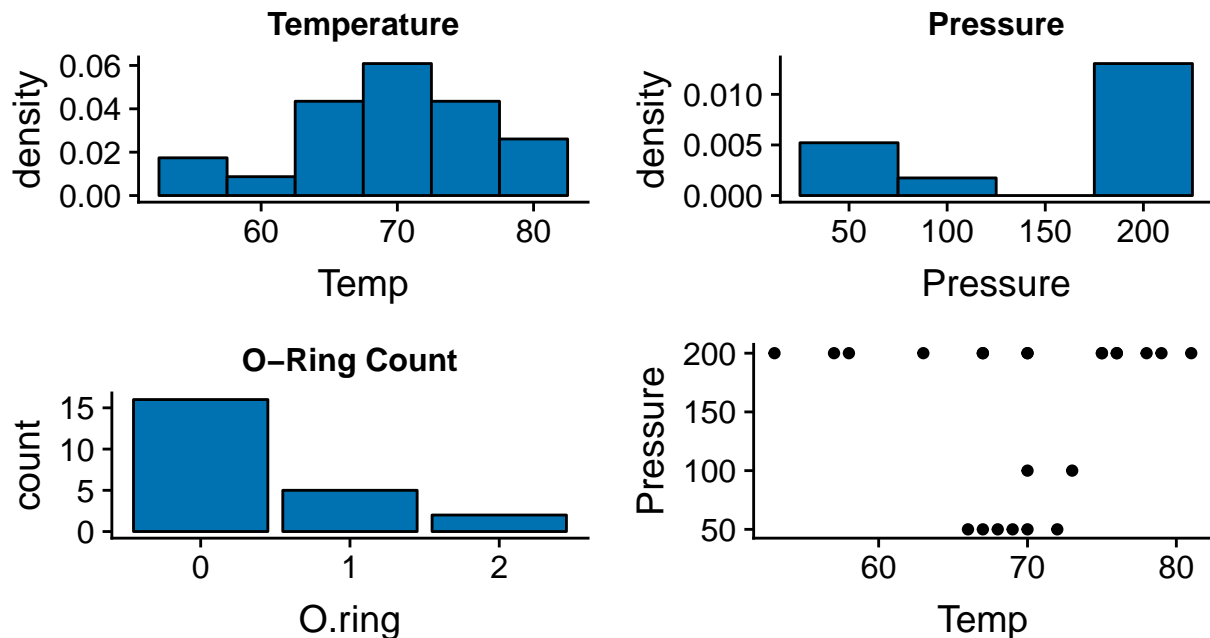

In the following plots, we examine the relationship between the response variable of O-ring count and each of the explanatory variables indepdently. In the O-ring Failures vs Temperature plot, we

see a clear negative relationship, where the number of O-rings are roughly decreasing as Temperature increases, except for a clear outlier at (Temp = 75, O.ring=2). Even the lowest temperature in cases with no incident is higher than the median temperature of cases with observed failures. On the other hand, we do not see as clear of a relationship between the O-rings failure count and Pressure, only that larger O-ring counts are associated with fewer Pressure values. In particular, there are no incidents of failure for a pressure of 100 psi, which does call into question the feasibility of including Pressure as an explanatory variable in modeling failure predictions.

```r
#temperature by o_ring
oring.temp = ggplot(data, aes(f_O.ring, Temp))+geom_boxplot(aes(fill = f_O.ring))+
  geom_jitter()+ggtitle("O-rings Failures by Temperature")+
  theme(plot.title = element_text(lineheight=1, face="bold",size = 10))
#pressure by o_ring
oring.pressure=ggplot(data, aes(f_O.ring, Pressure))+geom_boxplot(aes(fill = f_O.ring))+
  geom_jitter()+ggtitle("O-rings Failures by Pressure")+
  theme(plot.title = element_text(lineheight=1, face="bold",size = 10))
#all 3
all3=ggplot(data,aes(factor(f_pressure),Temp))+geom_boxplot(aes(fill = factor(f_O.ring)))+
  geom_jitter()+ggtitle("Damaged O-rings by Temperature and Pressure") +
  theme(plot.title = element_text(lineheight=1, face="bold",size = 10))
plot_grid(oring.temp, oring.pressure, scale = 1,label_size = 7)
```



```
all3
```

Damaged O–rings by Temperature and Pressure
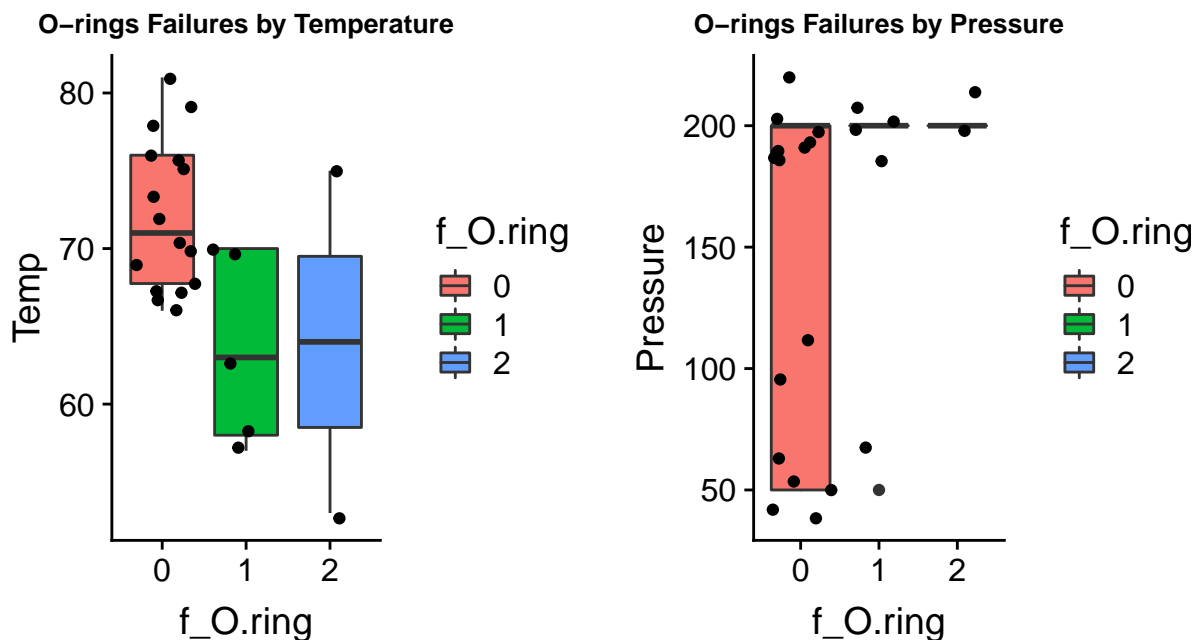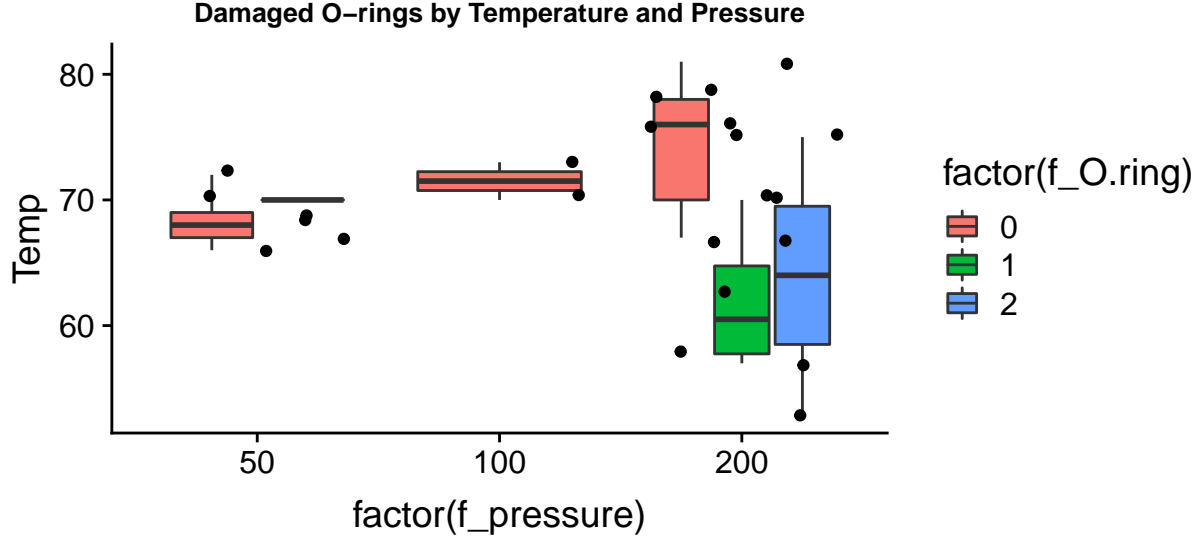
In the above plot, we attempt to ascertain relationships between all 3 variables by plotting them together. The data is partitioned by both O-ring failure count and Pressure, and we are able to see how both vary with Temperature. One thing we can see from this plot is that for cases of 0 O-ring failures, the relationship between temperature and pressure appear to be slightly positive, but this is not particularly useful for predicting cases where there is O-ring failure. More importantly, the relationship between Temperature and O-ring count becomes confounded when Pressure is used to partition the data. We see a positive relationship between Temperature and O-ring count for Pressure of 50 psi, which is counter to the observed relationship when we plotted only O-ring count versus Temperature. At 200 psi, there is no clear relationship between Temperature and O-ring count. This is further evidence that the inclusion of Pressure as an explanatory variable may not improve prediction results, as it may confound the relationship between Temperature and O-ring count.

4. Response Modeling with Temperature and Pressure Dependence

A. Importance of Independence Assumption

The logistic regression model used by Dalal et al to predict single O-ring failure has the following form:

$$logit(\hat{\pi}) = log(\frac{\hat{\pi}}{1 - \hat{\pi}}) = \hat{\beta}_0 + \hat{\beta}_1 \cdot Temp + \hat{\beta}_2 \cdot Pressure$$

where the logit transformation of the probability of failure, or the log odds of failure, is predicted by a linear component that involves Temperature and Pressure explanatory variables. The indpendence condition is necessary in predicting the actual response variable, which is the expected number of O-ring failures. Given estimated single O-ring failure $\hat{\pi}$, the probability of w O-rings failing out of a total of 6 is modeled by the binomial probability distribution:

$$p(w) = (nCw)\,(\hat{\pi})^w(1 - \hat{\pi})^{n-w}$$

where the expected number of failures is then calculated by

$$E[w] = \sum_1^n w \cdot p(w) = \sum_1^n w \cdot (nCw)(\hat{\pi})^w(1 - \hat{\pi})^{n-w} = \hat{\pi} \cdot n$$

4

In order to use binomial probability distribution to predict O-ring failure count, the assumptions of the binomial distribution must be met. The 5 assumptions are (1) all 6 O-rings are identical, (2) Each O-ring can only have one of 2 possible outcomes - failure or success, (3) The O-rings are independent from each other (4) the probabilities of success and failure are the same for each O-ring and (5) the random variable of interest is `w`, the number of O-rings that fail, and we are not interested in the order of failures.

In the actual context of the problem, 4 of the 5 assumptions are easily met by default. However assumption (3) is the only one that has potential problems as 3 O-rings are used to assemble each of the two rockets. The failure of one O-ring could lead to erosion of or blowby through the adjacent O-ring, thereby influencing the potential failure of O-rings within the same rocket. Therefore, the assumption that O-rings are independent from each other is not an entirely realistic representation of the mechanics of the rocket's operations.

To alleviate the author's concerns about independence, they also use an alternative model, which we will subsequently refer to as binary logistic regression model, to predict $\hat{\pi}^*$, which is defined as the probability of at least one O-ring failure. Since even a single O-ring failure could lead to catastrophic shuttle failure, this model can still be useful in predicting the chances of a disaster. In this model, the number of O-ring failures is transformed to a binary response, and the indpendence assumption of O-rings is not required. They can also compare the performance of the binary logistic regression model to the binomial logistic regression model in predicting the number of O-ring failures using inverse of the transformation

$$\hat{\pi}^* = 1 - (1 - \hat{\pi})^6$$

B. Logistic Regression Model Estimation using Generalized Linear Model

In the binomial logistic regression model below, we are specifically modeling the probability of single O-ring failure through this equation $logit(\hat{\pi}) = log(\frac{\hat{\pi}}{1 - \hat{\pi}}) = \hat{\beta}_0 + \hat{\beta}_1 \cdot Temp + \hat{\beta}_2 \cdot Pressure$ to find estimates for the coefficients of Temperature and Pressure. We transform the response varible into a proportion of `w/n` where the weights argument in the glm() function is set to n=6 to ensure that binomial probability distribution is used to model `p(w)` for `w` failures of 6 O-rings.

```
data[,proportion := O.ring/Number]
binomial.mod.glm.1=glm(proportion~Temp+Pressure,family=binomial,weights=Number,data=data)
summary(binomial.mod.glm.1)
```

```
##
## Call:
## glm(formula = proportion ~ Temp + Pressure, family = binomial,
##     data = data, weights = Number)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0361  -0.6434  -0.5308  -0.1625   2.3418
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.520195   3.486784   0.723   0.4698
## Temp        -0.098297   0.044890  -2.190   0.0285 *
```

5

```
## Pressure      0.008484    0.007677    1.105    0.2691
## ---
## 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: 16.546  on 20  degrees of freedom
## AIC: 36.106
##
## Number of Fisher Scoring iterations: 5
```

We see from this output that the paramters of the linear component are estimated to be $\hat{\beta}_0 = 2.5202$, $\hat{\beta}_1 = -0.0983$ and $\hat{\beta}_2 = 0.0085$, which closely match the results from Dalal et al.

On the other hand, the binary logistic regression model predicts $\hat{\pi}^*$, the probability of at least one O-ring failure using a similar equation - $logit(\hat{\pi}^*) = log(\frac{\hat{\pi}^*}{1 - \hat{\pi}^*}) = \hat{\beta}_0 + \hat{\beta}_1 \cdot Temp + \hat{\beta}_2 \cdot Pressure$. This model also uses glm() but the response variable undergoes the following transformation instead of proportions:

- O.ring $= 1$ or $2 \rightarrow$ O.ring.ALO $= 1$
- O.ring $= 0 \rightarrow$ O.ring.ALO $= 0$

where O.ring.ALO uses 1 to represent the existence of at least one failure. Furthermore, the weights argument, or the total number of O-rings, is no longer needed in constructing this model.

```
data[, O.ring.ALO := 0]
data[O.ring != 0]$O.ring.ALO = 1
binary.mod.glm.1=glm(O.ring.ALO~Temp+Pressure,family=binomial,weights=Number,data=data)
summary(binary.mod.glm.1)
```

```
##
## Call:
## glm(formula = O.ring.ALO ~ Temp + Pressure, family = binomial,
##     data = data, weights = Number)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.9377  -1.4152  -1.0402   0.8631   5.2538
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 13.292360   3.128750   4.248 2.15e-05 ***
## Temp        -0.228671   0.044902  -5.093 3.53e-07 ***
## Pressure     0.010400   0.003666   2.837  0.00455 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

6

```
##      Null deviance: 169.60  on 22  degrees of freedom
## Residual deviance: 112.69  on 20  degrees of freedom
## AIC: 118.69
##
## Number of Fisher Scoring iterations: 5
```

We see from this output that the paramters of the linear component are estimated to be $\hat{\beta}_0 = 13.2924$, $\hat{\beta}_1 = -0.2287$ and $\hat{\beta}_2 = 0.0104$.

C. Variable Importance using Likelihood Ratio Test

We use the Anova() function to perform the Likehood Ratio Test of each explanatory variable's importance in the prediction of the log-odds of probability of failure. For the test below, each of the LR χ^2 values are calculated from comparing the full model including both Temp and Pressure terms to a null model where the term of interest is excluded.

```
Anova(binomial.mod.glm.1, test = 'LR')
```

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

```
Anova(binary.mod.glm.1, test = 'LR')
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: O.ring.ALO
##          LR Chisq Df Pr(>Chisq)
## Temp       46.525  1  9.044e-12 ***
## Pressure    9.199  1   0.002422 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see from the results that the Pressure variable does not produce a statistiscally significant χ^2 statistic (p-value = 0.2145) in the binomial model, which means that we cannot reject the null hypothesis where the binomial model has the form - $log(\frac{\hat{\pi}_0}{1 - \hat{\pi}_0}) = \hat{\beta}_0 + \hat{\beta}_1 \cdot Temp$. Therefore, we conclude that the Pressure variable is not important in the binomial model.

However, when we run the same LRT on the binary model, we see that the χ^2 statistic for the Pressure variable is statistically significant, although with a p-value much larger than that of the Temperature χ^2 statistic. This means that we can reject the null hypothesis that the Pressure variable is not important in the binary model. However we did want to make a note here that in subsequent analysis, we will mainly focus on the binomial model, as a goodness of fit test below:

```
AIC(binomial.mod.glm.1)
```

```
## [1] 36.10589
```

```
AIC(binary.mod.glm.1)
```

```
## [1] 118.6926
```

shows that the binomial model with a lower AIC value produces a better fit than the binary model. We also know from Dalal et al that a G-test performed to compare both models also shows that the binomial model is a better fit.

D. Pressure Removal

The issue with the pressure variable is that we do not have enough observations of failures for all values of pressure. In particular, we have no observations for Pressure of 100 psi. As we saw in the EDA, the relationship between Temperature and Pressure in the absence of the Pressure variable is clear, but when the data is segmented by the pressure variable, relationships between explanatory variables and the reponse variable becomes confounded. The results of our LRT shows the pressure variable in the binomial model does not contribute to predicting O-ring failure in a statistically significant way. Therefore, we have sufficient justification to remove Pressure as an explanatory variable from the binomial model.

There is always a potential problem of loss of information with excluding any variable. It is possible that Pressure could have contributed to an accurate model had we acquired more data for 100 psi. Furthermore we wanted to note that Pressure is important to the binary model and in future analysis outside of the scope of this report, it may be interesting to compare the binary model with Pressure and Temperature to the binomial model with only Temperature in terms of their performance.

5. Response Modeling with only Temperature Dependence

A. Logistic Regression Model Estimation using Generalized Linear Model

In the second version of model estimation, we attempt to find the β parameters associated with the following model equations using glm():

1. Binomial: $logit(\hat{\pi}) = log(\frac{\hat{\pi}}{1 - \hat{\pi}}) = \hat{\beta}_0 + \hat{\beta}_1 \cdot Temp$

```
binomial.mod.glm.2=glm(proportion~Temp,family=binomial,weights=Number,data=data)
summary(binomial.mod.glm.2)
```

```
##
## Call:
## glm(formula = proportion ~ Temp, family = binomial, data = data,
##     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
```

where the parameters of the linear component are estimated to be $\hat{\beta}_0 = 5.0850$ and $\hat{\beta}_1 = -0.1156$.

2. Binary: $logit(\hat{\pi}^*) = log(\frac{\hat{\pi}^*}{1 - \hat{\pi}^*}) = \hat{\beta}_0 + \hat{\beta}_1 \cdot Temp$

```
binary.mod.glm.2=glm(O.ring.ALO~Temp,family=binomial,weights=Number,data = data)
summary(binary.mod.glm.2)
```

```
## 
## Call:
## glm(formula = O.ring.ALO ~ Temp, family = binomial, data = data,
##     weights = Number)
## 
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5992  -1.8647  -0.9266   1.1081   5.4318
## 
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 15.04290    3.01231   4.994 5.92e-07 ***
## Temp        -0.23216    0.04419  -5.254 1.49e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 169.60  on 22  degrees of freedom
## Residual deviance: 121.89  on 21  degrees of freedom
## AIC: 125.89
## 
## Number of Fisher Scoring iterations: 5
```

with parameter estimations $\hat{\beta}_0 = 15.0429$ and $\hat{\beta}_1 = -0.2322$. Both results match the Dalal et al paper closely. We note here that while Pressure does have significance in the binary model, the

subsequent analysis comparing the binary and binomial results will do so using the same linear component, which includes only the Temperature explanatory variable.

## B. Probability and Expectation vs Temperature

From each model, we use the expression $\dfrac{e^{\hat{\beta}_0 + \hat{\beta}_1 \cdot Temp}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 \cdot Temp}}$ to calculate π in the case of the binomial model and π^* in the case of the binary model. Given that the binomial model predicts the probability of any one O-ring failure, π, while the binary model estimates the probability of at least one O-ring failure, π^*, we use the transformation, $\pi = 1 - (1 - \pi^*)^{\frac{1}{6}}$, to compare the predicted π values between the two models. We also need this transformation in order to calculate and compare the expected number of failures using the function $E(w) = \sum_1^n w \cdot (nCw)(\pi)^w (1 - \pi)^{n-w}$, which depends on π instead of π^*. The results are plotted below for both models

```r
expectation = function(p){
  #failure range
  x = 1:6
  #probabilities
  probs = dbinom(x = x, size = 6, prob = p)
  #expectation
  expectation = sum(x*probs)
  return(expectation)
}
pi.from.p.star = function(pi.star){
  pi = 1-(1-pi.star)^(1/6)
  return(pi)
}
pi.star.from.pi = function(pi){
  pi.star = 1-(1-pi)^6
  return(pi.star)
}
x.seq = seq(31,81, by=0.01)
prob.table = matrix(data=NA,nrow=length(x.seq),ncol=7)
counter=1
for (x in x.seq){
  pi.binomial=exp(coef(binomial.mod.glm.2)['(Intercept)']+coef(binomial.mod.glm.2)['Temp']*x
      )/(1+exp(coef(binomial.mod.glm.2)['(Intercept)']+ coef(binomial.mod.glm.2)['Temp']*x))
  pi.binomial.star = pi.star.from.pi(pi.binomial)
  expectation.w.binomial = expectation(pi.binomial)
  pi.binary.star=exp(coef(binary.mod.glm.2)['(Intercept)']+coef(binary.mod.glm.2)['Temp']*x
        )/(1+exp(coef(binary.mod.glm.2)['(Intercept)']+coef(binary.mod.glm.2)['Temp']*x))
  pi.binary = pi.from.p.star(pi.binary.star)
  expectation.w.binary = expectation(pi.binary)
  prob.table[counter,] = c(x, pi.binomial,pi.binomial.star,expectation.w.binomial,
                        pi.binary.star,pi.binary,expectation.w.binary)
  counter = counter+1
}
par(mfrow=c(1,2))
```

```r
plot(x=prob.table[,1],y=prob.table[,2],ylim=c(0,1),
     main = "Probability of Failure vs Temperature", xlab='Temperature',
     ylab="Probability of Failure",type="l", col = 'red',cex.main=0.7)
lines(x=prob.table[,1],y=prob.table[,3],ylim=c(0,1),type="l",lty=2,col='red')
lines(x=prob.table[,1],y=prob.table[,5],ylim=c(0,1),type="l", lty=2, col='blue')
lines(x=prob.table[,1],y=prob.table[,6],ylim=c(0,1),type="l", col='blue')
legend(60,1,legend=c('binomial pi','binary pi','binomial pi*','binary pi*'),
       col=c('red','blue','red','blue'), lty=c(1,1,2,2), cex=0.6)
plot(data$Temp, data$O.ring, xlim = c(30, 82), ylim = c(0,6),
     main = "Expected Number of Failure vs Temperature",
     xlab='Temperature', ylab="Expected Number of Failure",cex.main=0.7)
lines(x=prob.table[,1],y=prob.table[,4],ylim=c(0,6),xlim=c(30, 82),type="l",col ='red')
lines(x=prob.table[,1],y=prob.table[,7],ylim=c(0,6),xlim=c(30, 82),type="l",col ='blue')
legend(60,6,legend=c('binomial','binary'),col=c('red','blue'),lty=c(1,1),cex=0.6)
```


Probability of Failure vs Temperature   **Expected Number of Failure vs Temperature**
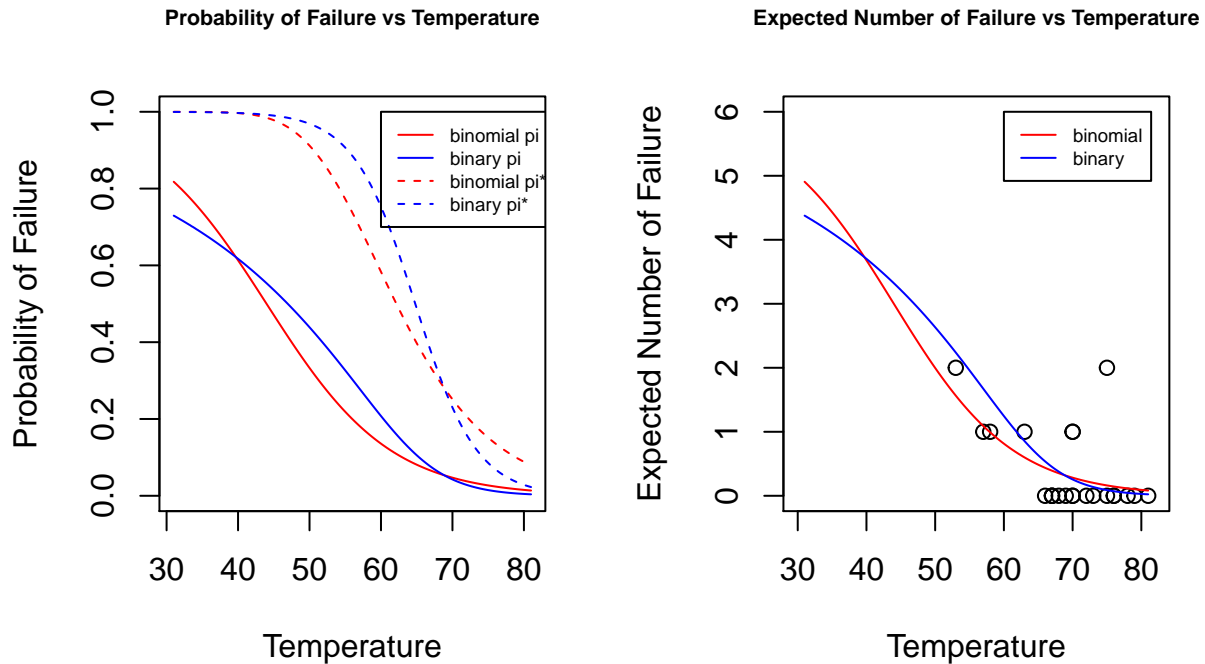
C. 95% Percent Wald Confidence Intervals

We use qnorm() to find the $Z_{\frac{\alpha}{2}}$ and $Z_{1-\frac{\alpha}{2}}$ associated with β_1 and the standard errors of the glm's linear component parameters to find the 95% confidence intervals for the β_1 parameter, under the assumption that β_1 is normally distributed. For the binomial model, we apply the inverse logit function $(f(x) = \dfrac{e^x}{1 + e^x})$ to the upper and lower bounds of the β parameters to find the upper and lower bounds of the 95% confidence intervals for $\hat{\pi}$. For the binary model, the process is similar, except an additional transformation is applied to transform from $\hat{\pi}^*$ to $\hat{\pi}$.

```r
predict.data = data.frame(Temp=(31:81))
Wald.CI.table = function(model, predict.data, alpha){
#logit(pi.hat) value
linear.pred=predict(object=model,newdata=predict.data,type="link",se=TRUE)
```

```r
#pi.hat.binomial
pi.hat =exp(linear.pred$fit)/(1+exp(linear.pred$fit))
#Confidence Interval
CI.lin.pred.upper=linear.pred$fit+qnorm(alpha/2)*linear.pred$se
CI.lin.pred.lower=linear.pred$fit+qnorm(1-alpha/2)*linear.pred$se
CI.pi.upper = exp(CI.lin.pred.upper)/(1+exp(CI.lin.pred.upper))
CI.pi.lower = exp(CI.lin.pred.lower)/(1+exp(CI.lin.pred.lower))
CI.table = data.frame(predict.data,pi.hat,CI.pi.upper,CI.pi.lower)}
#The CI table can be constructed directly from the function above for the binomial model
Binomial.Wald.CI.table=Wald.CI.table(model=binomial.mod.glm.2,
                                      predict.data=predict.data,alpha=0.05)
#However, for the binary model, the function produces pi.hat.star and the CI for pi.hat.star
#We will need to apply the pi.from.p.star transformation to the resulting values
Binary.Wald.CI.table.star=Wald.CI.table(model=binary.mod.glm.2,
                                         predict.data=predict.data,alpha=0.05)
Binary.Wald.CI.table = data.frame(Temp=Binary.Wald.CI.table.star[,1],
                      pi.hat=sapply(Binary.Wald.CI.table.star[,2], pi.from.p.star),
                      CI.pi.upper=sapply(Binary.Wald.CI.table.star[,3], pi.from.p.star),
                      CI.pi.lower=sapply(Binary.Wald.CI.table.star[,4], pi.from.p.star))
par(mfrow=c(1,2))
plot(x=Binomial.Wald.CI.table[,1],y=Binomial.Wald.CI.table[,2], ylim=c(0,1),
     main = "Probability of Failure vs Temperature (Binomial Model)",cex.main=0.7,
     xlab='Temperature',ylab="Probability of Failure",type="l", col = 'red')
lines(x=Binomial.Wald.CI.table[,1],y=Binomial.Wald.CI.table[,3],ylim=c(0,1),
     type="l", lty = 2, col = 'pink')
lines(x=Binomial.Wald.CI.table[,1],y=Binomial.Wald.CI.table[,4],ylim=c(0,1),
     type="l", lty = 2, col = 'pink')
plot(x=Binary.Wald.CI.table[,1],y=Binary.Wald.CI.table[,2],ylim=c(0,1),
     main = "Probability of Failure vs Temperature (Binary Model)",cex.main=0.7,
     xlab='Temperature',ylab="Probability of Failure",type="l",col='blue')
lines(x=Binary.Wald.CI.table[,1],y=Binary.Wald.CI.table[,3],ylim=c(0,1),
     type="l", lty = 2, col = 'light blue')
lines(x=Binary.Wald.CI.table[,1],y=Binary.Wald.CI.table[,4],ylim=c(0,1),
     type="l", lty = 2, col = 'light blue')
```

Probability of Failure vs Temperature (Binomial Model)   **Probability of Failure vs Temperature (Binary Model)**
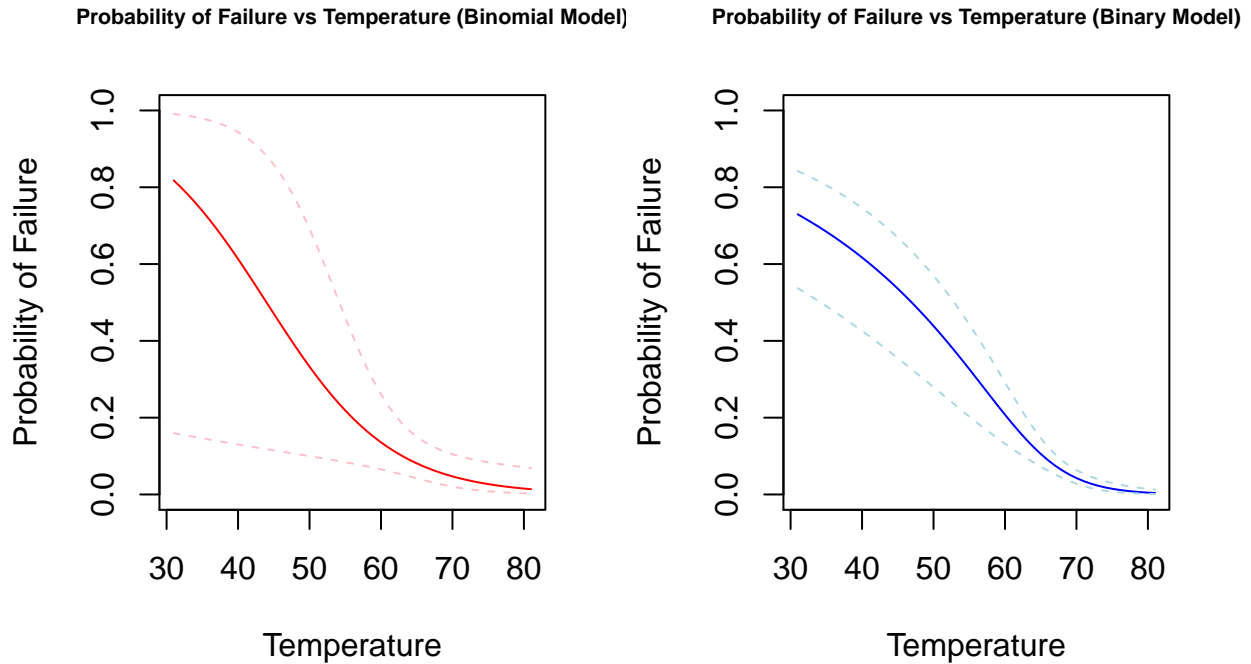
Given the order of transformations applied to determine the 95% confidence intervals for the binary model, it appears that the confidence interval for the binary model is narrower than that of the the binomial model. For both models, the confidence interval is much wider for lower temperatures than that of higher temperatures. This is due to the lack of observable data for temperatures less than 51°, as most of our data points used to fit either model are associated with higher temperatures. Given that the density of points decrease as the temperature decreases, the prediction of $\hat{\pi}$ for lower temperatures become increasing less reliable, leading to wider 95% confidence intervals.

D. O-Ring Estimation Failure at 31 Degrees

We performed a second set of AIC calculations to measure goodness of fit between the binomial and binary models that use Temperature as only explanatory variable. Again, we find below that the binomial model outperforms the binary model with a better fit and lower AIC. Our subsequent analysis for this section and remaining sections will be exclusive to the binomial model.

```
AIC(binomial.mod.glm.2)
```

```
## [1] 35.64654
```

```
AIC(binary.mod.glm.2)
```

```
## [1] 125.8912
```

We can pull the estimated probability of failure and its 95% Wald confidence interval from the table constructed in part C:

```
Binomial.Wald.CI.table[Binomial.Wald.CI.table$Temp == 31,]
```

```
##   Temp    pi.hat CI.pi.upper CI.pi.lower
## 1   31 0.8177744   0.1596025   0.9906582
```

The Wald confidence interval assumes that the parameters β_0, the intercept of the linear component,

and β_1, the parameter associated with log-odds change per change in temperature, are normally distributed. Values for either parameter that are more than ~2 standard errors away from the estimated values are highly unlikely.

However, given that the Wald Confidence interval usually has a confidence level that is lower than the stated confidence level due to the discreteness of the binomial response variable distribution, we also calculate the 95% profile likelihood CI with true confidence level closer to 95%:

```r
point = matrix(data = c(1, 31), nrow = 1, ncol = 2)
#mcprofile used to find profile likelihood CI:
profile.LR.CI = function(model, testpoint){
linear.combo <- mcprofile(object = model, CM = testpoint)
ci.logit.profile <- confint(object = linear.combo, level = 0.95)
pi.hat = exp(ci.logit.profile$estimate)/(1+exp(ci.logit.profile$estimate))
pi.hat.int = exp(ci.logit.profile$confint)/(1+exp(ci.logit.profile$confint))
data.frame(pi.hat,lower = pi.hat.int[1],upper=pi.hat.int[2])}
(Binomial.PLR.CI = profile.LR.CI(model = binomial.mod.glm.2, testpoint = point))
```

```
##    Estimate    lower     upper
## C1 0.8177744 0.1418508 0.9905217
```

We see that the 95% confidence intervals are quite similar using either method, with Profile Likelihood giving a lower value for the lower bound. The estimated probability of single O-ring failure is 0.8178, and the lowest bound from both intervals is 0.1419. This value is quite close to 1/6, the minimum probability that is needed for a calculated expectation of exactly 1 O-ring failure. This means that even with a very wide confidence interval for lower values of temperature, the expectation of one O-ring failure at 31° is still highly probable, given the average expected O-ring failure count at this temperature to be 4 with a 95% confidence interval that is bounded by approximately 1 at the lower end. This means that catastrophic failure on the day of the Challenger launch, during which the temperature was 31°, was likely inevitable.

E. Bootstrap method for 90% confidence intervals

We simulate the bootstrap method from Dalal et All to produce 10,000 binomial logistic regression models. In the code below, we construct each trial by first randomly sampling 23 Temperatures and predicting their O-ring count response using rbinom(), where the probability of single O-ring failure is predicted by the model parameters in part 5A. Given the distribution shape of the Temperature variable we found in EDA, we use a t-distribution with mean and standard deviation calculated from the observed temperature data to sample Temperatures for each trial. We individually fit a binomial logistic regression model to each of the k=23 sample points and this is repeated to find 10,000 pairs of estimated parameters (β_0, β_1).

For the plots of points that estimate the 90% confidence interval, we use the 10,000 models to predict the probability of single O-ring failure for a series of temperature values between 31° and 81°. For each temperature value, we take the 5th percentile and 95th percentile values from 10,000 predictions to construct the bands of the 90% confidence interval. The results are plotted to check the shape of our confidence intervals -

```r
#gather estimated beta from 10000 samples
b.seq = seq(1,10000, by=1)
```

```r
simulated.data = function(beta){
  #sample temp from t-distribution
  temp = rt(23, df = 22)*sd(data$Temp)+mean(data$Temp)
  pi = exp(beta['(Intercept)']+beta['Temp']*temp)/(
    1+exp(beta['(Intercept)']+beta['Temp']*temp))
  sim.O.ring = rbinom(n=length(temp), size = 6, prob = pi)
  sim.mod.glm = glm(sim.O.ring/6 ~ temp, family=binomial,
                    weights=rep(6, length(temp)), control = list(maxit = 100))
  return(c(coef(sim.mod.glm)['(Intercept)'], coef(sim.mod.glm)['temp']))}
bootstrap.betas = matrix(data=NA,nrow=length(b.seq),ncol=2)
counter=1
for (b in b.seq){
  bootstrap.betas[counter, ]=simulated.data(beta=coef(binomial.mod.glm.2))
  counter = counter + 1}
get_conf_bounds = function(bootstrap.betas, temp){
  predictions = matrix(data=NA,nrow=length(b.seq),ncol=1)
  c=1
  for(i in 1:length(bootstrap.betas[,1])){
    predictions[c] = exp(bootstrap.betas[i,][1] + bootstrap.betas[i,][2]*temp)/(
      1+exp(bootstrap.betas[i,][1] + bootstrap.betas[i,][2]*temp))
    c = c + 1}
  return(quantile(sort(predictions), probs = c(0.05, 0.95), na.rm = TRUE))}
```

```r
t = seq(30,82,5)
conf_bounds = sapply(t, function(t){get_conf_bounds(bootstrap.betas,t)})
(conf_bounds31 = get_conf_bounds(bootstrap.betas, 31))
```

```
##        5%        95%
## 0.1302853 0.9955238
```

```r
(conf_bounds72 = get_conf_bounds(bootstrap.betas, 72))
```

```
##          5%          95%
## 0.009688862 0.069551834
```

```r
plot(x=Binomial.Wald.CI.table[,1],y=Binomial.Wald.CI.table[,2],xlim=c(30,82),ylim = c(0,1),
     main = "Probability of Failure vs Temperature (Binomial Model) with 90% Bootstrap CI",
     xlab='Temperature',ylab="Probability of Failure",col='red',type ='l',cex.main=0.9)
lines(x=t, y = conf_bounds[1,], ylim=c(0,1), type='p', col = 'pink')
lines(x=t, y = conf_bounds[2,], ylim=c(0,1), type='p', col = 'pink')
points(c(31,31),conf_bounds31, col = 'orange')
points(c(72,72),conf_bounds72, col = 'orange')
```

Probability of Failure vs Temperature (Binomial Model) with 90% Bootstrap CI
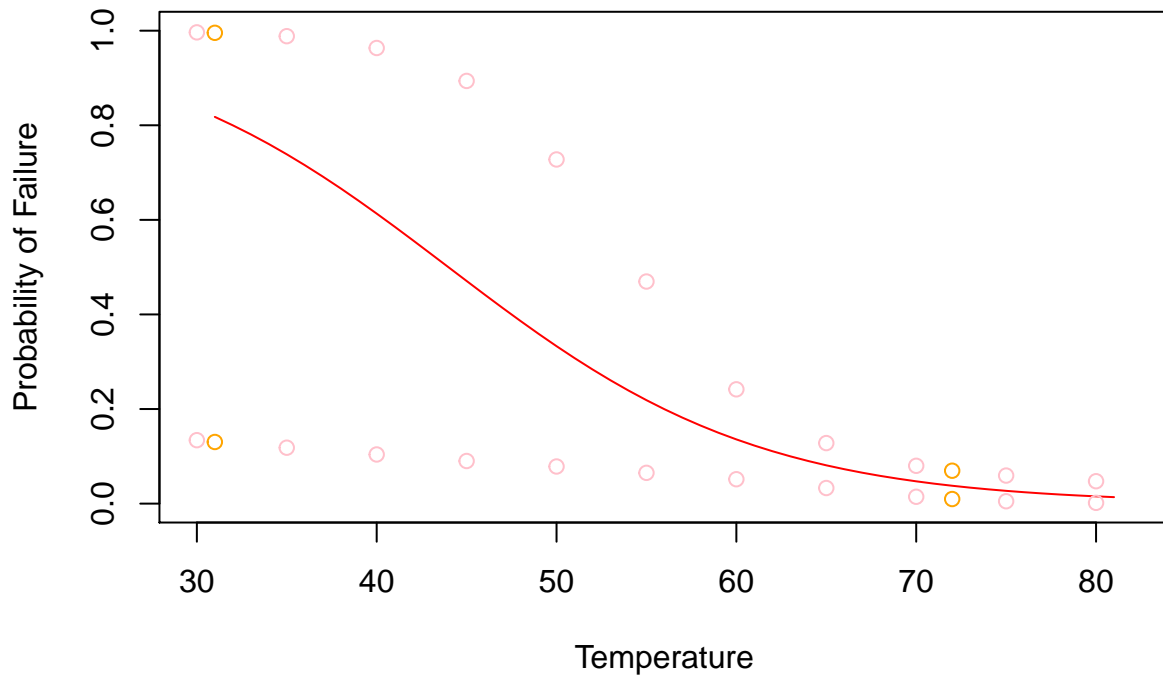
Using the bootstrap method, the 90% confidence intervals at 31° and 72° are (0.1319, 0.9964) and (0.0091, 0.0704), respectively and are plotted as the 4 orange points above. The lower bound for the bootstrap 90% confidence interval at 31° is slightly lower than either of the 95% confidence intervals from Wald and profile likelihood calculated in the previous section. This suggest that the bootstrap method leads to a more conservative estimate, or a wider interval, for the confidence interval of the same level. Similar to the confidence bounds we generated in 5B, we can observe that the bounds are tight around our model for high temperatures, then expand as the temperature gets lower. This is caused by the distribution of our temperature variable, since we attempted to preserve the distribution of temperatures when sampling from a t-distribution constructed from the observed temperature data leading to a range of temperature similar to the original dataset for each k=23 trial. Even though we have many more observations in this method, we can only be more certain about simulated observations that look similar to actually observed data. Since we have much fewer real observations with low temperatures, we would be less certain about predictions for simulated low temperatures, hence the wider confidence intervals for lower temperatures.

F. Quadratic Dependence on Temperature

```
binomial.mod.glm.temp_squared=glm(formula=proportion ~Temp +I(Temp^2),family=binomial,
                           weights=Number,data=data)
anova(binomial.mod.glm.2,binomial.mod.glm.temp_squared,test="Chisq")

## Analysis of Deviance Table
##
## Model 1: proportion ~ Temp
## Model 2: proportion ~ Temp + I(Temp^2)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
```

```
## 1        21     18.086
## 2        20     17.592  1   0.4947   0.4818
```

We use anova() function to perform the Likelihood Ratio Test in order to determine the importance of the $Temp^2$ term. In this test, we compare the new model $logit(\hat{\pi}) = log(\dfrac{\hat{\pi}}{1 - \hat{\pi}}) = \hat{\beta}_0 + \hat{\beta}_1 \cdot Temp + \hat{\beta}_2 \cdot Temp^2$ to the null model $logit(\hat{\pi}_0) = log(\dfrac{\hat{\pi}_0}{1 - \hat{\pi}_0}) = \hat{\beta}_0 + \hat{\beta}_1 \cdot Temp$ to obtain the χ^2 statistic. From here, we see that the χ^2 statistic produced a statistically insignificant p-value of 0.4818, which is insufficient for us to reject the null hypothesis that the $Temp^2$ term is not important. Therefore, our final model will only include the linear $Temp$ term and exclude the $Temp^2$ term.

Results Interpretation and and Comparison to Linear Model

A. Final Model Results Interpretation

Our final model from the previous section is the binary logistic regression model containing only Temperature as a explanatory variable ($logit(\hat{\pi}) = log(\dfrac{\hat{\pi}}{1 - \hat{\pi}}) = \hat{\beta}_0 + \hat{\beta}_1 \cdot Temp$). We used this model to predict the probability of any one O-ring failure on the day of the shuttle launch, where Temperature = 31°, to be approximately 0.8178 with confidence intervals calculated from 3 different methods:

Method	Confidence Interval	Confidence Level
Wald	(0.1596, 0.9907)	95%
Profile Likelihood	(0.1419, 0.9905)	95%
Bootstrap Method	(0.1319, 0.9964)	90%

Given that we were only able to fit the binomial logistic regression model using Temperatures above 51°, the estimate for the launch day failure probability is extrapolated and all 3 confidence intervals are relatively wide. However, as previously discussed, the average predicted expected number of O-ring failures given the predicted probability is 4, with a minimum of 1, since the lower bounds of the confidence intervals for $\hat{\pi}$ are all fairly close to the probability of 1/6, the value of π needed to result in an expected value of 1 O-ring failure out of 6 under the binomial distribution. Therefore, we conclude from these results that the final model would have predicted with high certainty that one O-ring would have failed on the day of the launch at temperature 31°, leading to potential catastrophic failure of the suttle.

The odds of O-ring failure on the day of the Challenger launch is defined as the ratio of the probability of failure to probability of success (O-ring does not fail). This is calculated here by exponentiating the linear component

```
(odds = exp(coef(binomial.mod.glm.2)['(Intercept)'] + coef(binomial.mod.glm.2)['Temp']*31))
```

```
## (Intercept)
##    4.487703
```

We see from this result the probability that a single O-ring fails is approximately 4.4877 times that of the probability that a single O-ring does not fail.

Given that the final binary logistic regression model contains only one explanatory variable, we can estimate the change in the odds-ratio through the following equation -

$$\widehat{OR} = \frac{Odds_{Temp+c}}{Odds_{Temp}} = exp(c\hat{\beta}_1)$$

We can see from this equation that the change in the odds-ratio of single O-ring failure probability does not depend on the value of Temperature, but only the change in Temperature. Given the distribution of observed Temperature values, we calculate the odds-ratio change for 1 standard deviation of 7.0571 degrees in Temperature change below:

```
c = sd(data$Temp)
odds.ratio = exp(coef(binomial.mod.glm.2)['Temp']*c)
(1/odds.ratio)
```

```
##      Temp
## 2.260999
```

We then take the inverse of the odds-ratio calculation, 1/odds-ratio, to specifically measure the change in the odds ratio as the temperature decreases. From this result, we have that given 1 standard deviation of temperature decrease of 7.0571 degrees, the odds of O-ring failure becomes 2.261 times as large.

B. Comparison with Linear Regression Model

```
model.linear.2 = lm(O.ring ~ Temp, data = data)
par(mfrow=c(2,2))
plot(model.linear.2, which=1, cex.main = 0.7, cex.lab = 0.7, cex.axis = 0.7)
plot(model.linear.2, which=2, cex.main = 0.7, cex.lab = 0.7, cex.axis = 0.7)
plot(model.linear.2, which=3, cex.main = 0.7, cex.lab = 0.7, cex.axis = 0.7)
plot(model.linear.2, which=5, cex.main = 0.7, cex.lab = 0.7, cex.axis = 0.7)
```
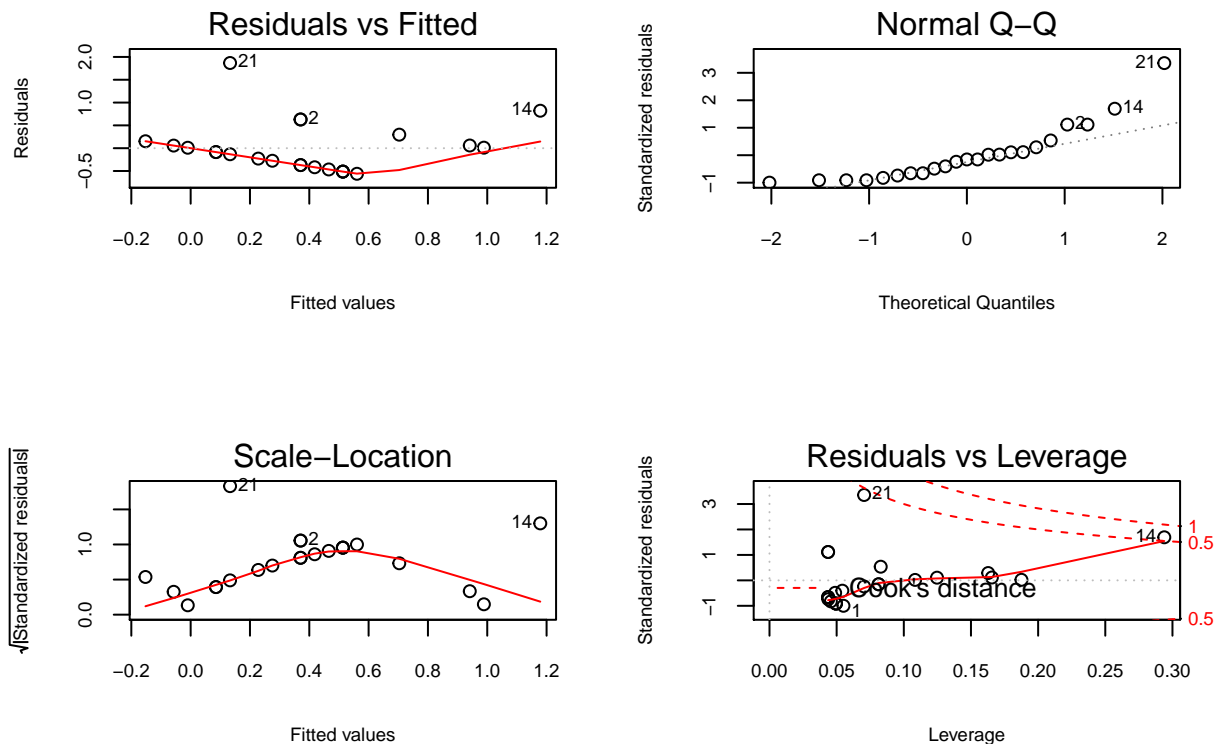
When we use linear regression we need to meet several conditions:

1. Each O-ring failure must be independent. No O-ring failure should have any impact on another O-ring failure.
2. There must exist a linear relationship between the temperature and the number of failed O-rings.
3. The variance in the number of O-rings must remain constant across temperatures.
4. The errors must be normally distributed.

In addiion we must also check that the assumptions of the Classical Linear Model hold. We used our diagnostic plots to examine some these conditions.

1. The Residuals vs Fitted plot shows us that here are still patterns in the residuals. This is due to the fact that we are predicting discrete values rather than continuous.
2. The Normal Q-Q plot shows us that with the exception of a few points, the residuals are somewhat normally distributed.
3. The scale location plot shows us that the variance is not completely stable.
4. The Residuals vs Leverage plot shows us that there are two points with Cook's distance near 0.5, indicating that they have some leverage. However, since there are no points with Cook's distance greater than 1, we can say that no point has high leverage.

Given that we have clearly violated at least one of the assumptions of the Classical Linear Model, in particular Homoskedasticity as seen in the Scale-Location plot, we do need to note that the estimators from fitting an Ordinary Least Squares linear regression model on O-ring failures versus Temperature may not be entirely unbiased. However, we will continue here with the comparison of the linear model to the linear component of the binary logistic regression model.

```
stargazer(binomial.mod.glm.2, model.linear.2, type = "latex")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu % Date and time: Fri, Sep 27, 2019 - 18:22:55

Table 2:

	Dependent variable:	
	proportion	O.ring
	logistic	*OLS*
	(1)	(2)
Temp	−0.116**	−0.048**
	(0.047)	(0.017)
Constant	5.085*	3.698***
	(3.052)	(1.220)
Observations	23	23
R^2		0.261
Adjusted R^2		0.226
Log Likelihood	−15.823	
Akaike Inf. Crit.	35.647	
Residual Std. Error		0.577 (df = 21)
F Statistic		7.426** (df = 1; 21)

Note: *p<0.1; **p<0.05; ***p<0.01

We calculate the predicted number of O-ring falures on the day of the Challenger launch at Temperature of 31° and its 95% Wald Conifdence interval Using the standard error of β_1 from the linear model below.

```
Linear.Pred.31 = model.linear.2$coefficients[1]+model.linear.2$coefficients[2]*31
Linear.Model.SE = 0.017
CI.lin.pred.upper=Linear.Pred.31+qnorm(0.05/2)*Linear.Model.SE
CI.lin.pred.lower=Linear.Pred.31+qnorm(1-0.05/2)*Linear.Model.SE
data.frame(estimate = Linear.Pred.31, Upper = CI.lin.pred.upper, Lower = CI.lin.pred.lower)
```

```
##             estimate    Upper    Lower
## (Intercept) 2.224683 2.191363 2.258002
```

We see that while the binomial logistic regression function predicts the expected number of O-ring failure to be approximately 4, with 95% confidence interval of approximately (1,6), the linear model predicts the expected number of O-ring falilure to be approximately 2, with a much narrower 95% confidence interval of (2.1913, 2.2580). Had we used a linear model to fit the observed data and extrapolated to 31° we still would have predicted that at least one O-ring failure would have occurred in the Challenger on the day of the launch, leading to potential catastrophic failure.

A challenge with comparing the goodness of fit between the models is that they are different types of models and therefore, the coefficients β_0 and β_1 cannot simply be compared equally between the two. To reconcile this issue, we can compare the sum of squared errors, calculated from the difference between the fitted and expected values, between the two models.

```
#sum of square errors
paste("Binomial SSE: ", sum((binomial.mod.glm.2$fitted.values*6 - data$O.ring)^2))
```

```
## [1] "Binomial SSE:  5.99202487479843"
```

```
paste("Linear SSE: ", sum((model.linear.2$fitted.values-data$O.ring)^2))
```

```
## [1] "Linear SSE:  7.00206349206349"
```

Using sum of squared error for the expected values, we can see that the binomial logistic regression model with SSE value of 5.9920 outperforms the linear model with SSE value of 7.0021. We plot a comparison of both models in predicting the expected number of O-ring failures below:

```
prob.table.linear = matrix(data=NA,nrow=length(x.seq),ncol=2)
counter = 0
for (x in x.seq){
  expectation.w.linear = coef(model.linear.2)['(Intercept)'] +
                          coef(model.linear.2)['Temp']*x
  prob.table.linear[counter,] = c(x, expectation.w.linear)
  counter = counter+1}
plot(data$Temp, data$O.ring, xlim = c(30, 82), ylim = c(0,6),
    main = "Expected Number of Failure vs Temperature",
    xlab='Temperature', ylab="Expected Number of Failure")
lines(x=prob.table[,1],y=prob.table[,4],ylim=c(0,6),xlim=c(30, 82),
     type="l", col = 'red')
lines(x=prob.table.linear[,1],y=prob.table.linear[,2],ylim=c(0,6),
     xlim=c(30, 82),type="l", col = 'green')
legend(70, 6, legend=c('binomial', 'binary', 'linear'),
       col=c('red', 'blue','green'), lty=c(1,1,1), cex=0.8)
```


Expected Number of Failure vs Temperature
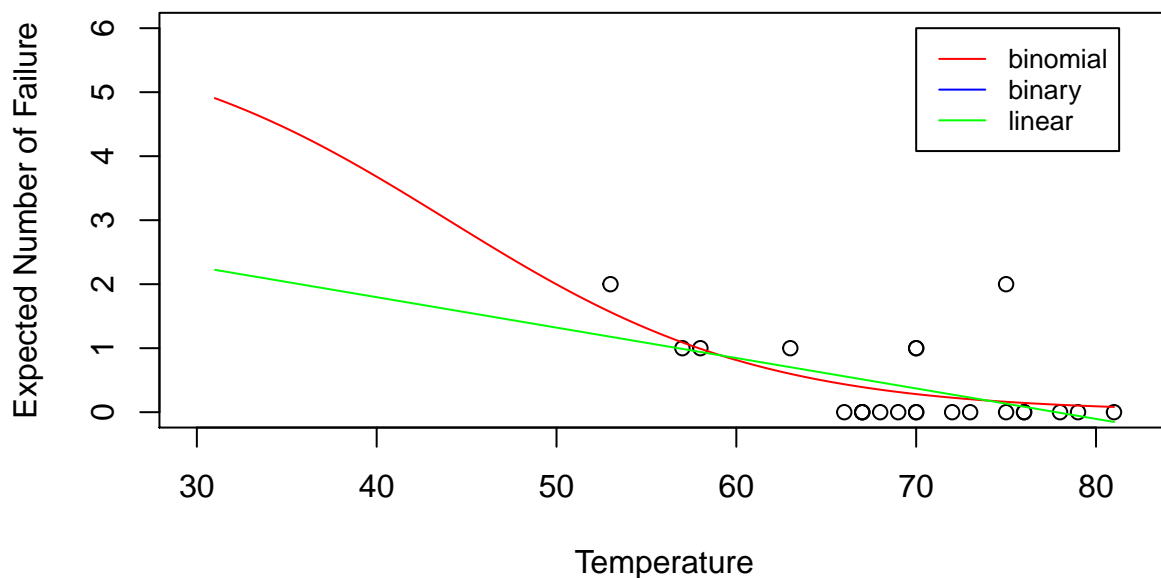
We can see from the plots of both models that while the linear model fits the points well for low

O-ring counts and high temperatures, the binomial logistic regression model clearly fits the data better for the same set of points. Furthermore, the binary logistic regression model makes more intuitive sense, as continously decreasing temperatures will lead to maximum expectation of 6 O-ring failures while increase in temperature will lead to minimum expectation of 0 O-ring failures. The linear model may fit a select sample of the data well, but it makes no sense for very low temperature or very high temperatures where the predicted number of O-ring failures may exceed the maximum of 6 or decrease past the minimum of 0, leading to a negative count for temperatures greater than 80°

Conclusion

In summary, our analysis of comparisons in performance of the 3 different models - binomial logistic regression, binary logistic regression , and linear regression - lead us to conclude one O-ring failure on the day of the Challenger launch was high probable. Although all 3 models could have predicted the Challenger disaster, we found that the binomial logistic regression model of the form $(logit(\hat{\pi}) = log(\frac{\hat{\pi}}{1 - \hat{\pi}}) = \hat{\beta}_0 + \hat{\beta}_1 \cdot Temp)$, where $\hat{\pi}$ is the probability of single O-ring failure, has the best overall fit and is therefore the most reliable model in predicting single O-ring failure. This model predicted an expected number of 4 O-ring failures at Temperature $= 31°$ with 95% confidence interval have lower bound close to 1 and upper bound close to 6.