W271 Group Lab 1: Investigation of the 1989 Space Shuttle Challenger Accident

Karthik Srinivasan, Anthony Ramirez, Kevin Fu

Introduction

In this lab, we analyze and test a variety of models to find a good predictor for o-ring failure. We determined that the most suitable model for this exercise is the logistic regression model as opposed to a linear regression model. Analysis via logistic regression show that there is strong statistical evidence of a temperature effect on incidents of o-ring thermal distress.

Part 1 (25 points) We import mandatory libraries and `challenger.csv` in order to perform EDA on the dataset. The library `GridExtra` is used to be able to put 2 graphs side by side in the report so that we don't take up more than 20 pages.

```
list.of.packages <-
  c("car", "dplyr", "Hmisc", "skimr", "ggplot2", "stargazer", "mcprofile", "hrbrthemes")
library(car)
library(Hmisc)
library(skimr)
library(ggplot2)
library(stargazer)
library(mcprofile)
library(dplyr)
library(gridExtra)

df <- read.csv(file = "challenger.csv", header = TRUE, sep = ",")
df <- df[c(2:5)]
skim(df)
```

Table 1: Data summary

Name	df
Number of rows	23
Number of columns	4
Column type frequency:	
numeric	4
Group variables	None

Variable type: numeric

| skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Temp	0	1	69.57	7.06	53	67	70	75	81	
Pressure	0	1	152.17	68.22	50	75	200	200	200	
O.ring	0	1	0.39	0.66	0	0	0	1	2	
Number	0	1	6.00	0.00	6	6	6	6	6	

We have a very small but clean dataset. Per above, there are no missing variables and no evidence of invalid/NA values.

We have 23 observations of total launches, in which 5 pieces of data (Flight, Temperature, Pressure, O.ring, Number) are recorded for each of those 23 launches.

On initial glance, there are 2 variables that do not provide much value other than serving as identifiers: `Flight`, which just tracks which flight number, 1 through 23, and `Number`, which is the value 6 across all 23 observations. We will use the variable `Number` in the following parts, but since we will not use the variable `Flight`, we have eliminated the variable in our dataset.

Therefore, we will spend most of our efforts on analyzing Temperature, Pressure, and Oring.

```
hist_temp <- ggplot(data = df, aes(Temp)) +
  geom_histogram(binwidth=1) + xlab("Temperature") + ylab("Instances") +
  ggtitle("Number of Flights at Certain Temperatures") +
  theme(plot.title = element_text(size=9))
hist_pressure <- ggplot(data = df, aes(Pressure)) +
  geom_histogram(binwidth=3) + xlab("Pressure") + ylab("Instances") +
  ggtitle("Number of Flights at Different PSI") +
  theme(plot.title = element_text(size=10))
grid.arrange(hist_temp, hist_pressure, ncol=2)
```
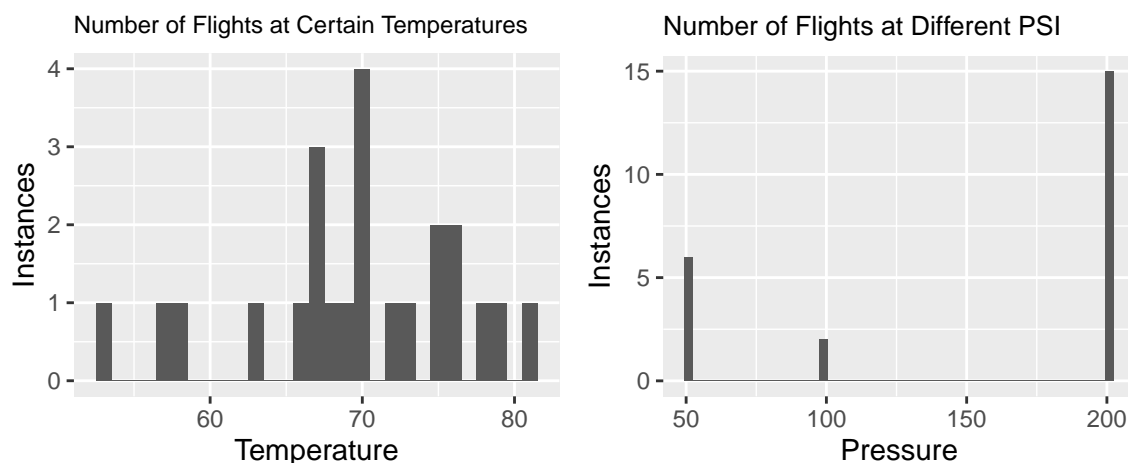

Temperature

- Temperatures are centered around 70 degrees Fahrenheit (as evident by the mean and median summary measures), and ranges from 53-81F.

- Only 4 temperatures appear more than once. 70 appears 4 times, followed by 67 appearing 3 times, followed by 75 and 76 degrees appearing twice.

```
table(df$Temp)
```

```
##
## 53 57 58 63 66 67 68 69 70 72 73 75 76 78 79 81
##  1  1  1  1  1  3  1  1  4  1  1  2  2  1  1  1
```

- 6 out of 7 failed launches (85.7%) occurred at or below 70F. With the data that we have, all launches (100%) had 1 or more o-ring failures if the temperature was below 64F. However, there is an exception/outlier where a failure occurred at 75F. Nevertheless, based on the box-plot below, we can clearly see that there is a higher chance of o-ring failure if the temperature was below 70F.

```
df_failures_only <- df %>% filter(df$O.ring > 0)
table(df_failures_only$Temp)
```

```
##
## 53 57 58 63 70 75
##  1  1  1  1  2  1
```

```
oring_temp <- ggplot(df, aes(factor(O.ring), Temp, color=)) +
  geom_boxplot(aes(fill = factor(O.ring))) + geom_jitter() +
  scale_fill_manual(values=c("#999999", "#E69F00", "#56B4E9")) +
  ggtitle("Temp at Various O rings (All Trials)") + xlab("O ring(s)") +
  theme(plot.title = element_text(size = 10, lineheight=1), legend.position = "none")

grid.arrange(oring_temp, ncol=1)
```
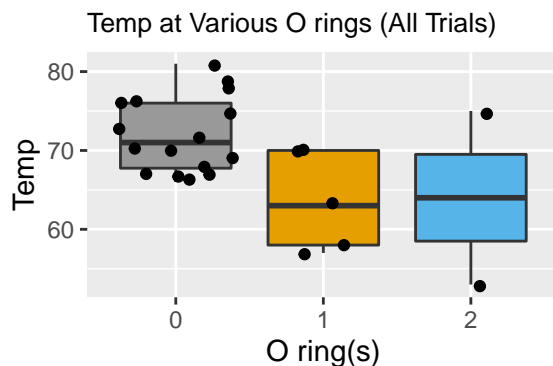

- Looking at the 7 failed launches only, the mean/median temperature was around 63-64F, which is exactly 1 standard deviation (7.06 degrees) below the mean/median temperature of 70F in the overall dataset.

Pressure

- We have 3 different pressure levels: 50, 100, and 200 psi. The minimum pressure level required for a successful launch is 50 psi. A majority (65.2%, 15 out of 23) of the launches were launched at the 200 psi level.

- Out of 15 total attempts at the 200 psi level, 9 were successful (60%) and 6 (40%) were "failures" (defined as 1 or more o-ring failures). However, 6 out of the 7 total failed launches (85.7%) occurred at the 200 psi level, and only 1 out of the 7 occurred at the 50 psi level.

```
table(df$Pressure)
```

```
##
##  50 100 200
##   6   2  15
```

```
table(df_failures_only$Pressure)
```

```
##
##  50 200
##   1   6
```

- We can see in this boxplot below that most failures in the 200 psi level mostly occurred below the 70 degree temperature mark.

```
pressure_temp <- ggplot(df, aes(factor(Pressure), Temp, color=)) +
  geom_boxplot(aes(fill = factor(Pressure))) + geom_jitter() +
  scale_fill_manual(values=c("#999999", "#E69F00", "#56B4E9")) +
  ggtitle("Temp at Various Psi (All Trials)") + xlab("Pressure") +
  theme(plot.title = element_text(size = 10, lineheight=1), legend.position = "none")
pressure_temp_failures <- ggplot(df_failures_only, aes(factor(Pressure), Temp, color=)) +
  geom_boxplot(aes(fill = factor(Pressure))) + geom_jitter() +
  scale_fill_manual(values=c("#999999", "#56B4E9")) +
  ggtitle("Temp at Various Psi (Failures Only)") + xlab("Pressure") +
  theme(plot.title = element_text(size = 10, lineheight=1), legend.position = "none")
grid.arrange(pressure_temp, pressure_temp_failures, ncol=2)
```
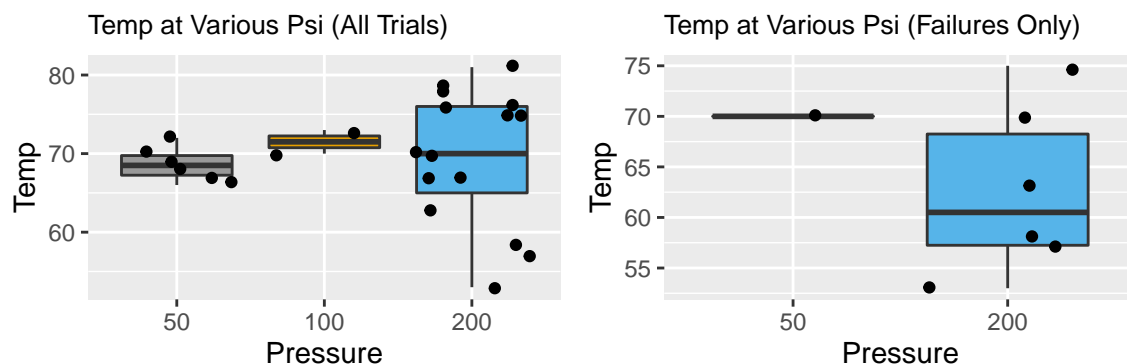

- Looking at the 7 failed launches only, the median pressure level remained the same at 200 psi, while the mean pressure levels increased by less than 1 standard deviation. Therefore, we do not have much significant evidence that 200 psi was a factor in contributing to O ring failures.

```
dat <- table(df$Pressure, df$O.ring)
barplot(as.matrix(dat), main="O-ring failures by PSI level",
  xlab="# of O-ring Failures", ylab="Total Instances", legend=T,
  legend.text = c("50 psi", "100 psi", "200 psi"),
```

```
  col=c("#999999", "#E69F00", "#56B4E9", cex.lab = 0.5))
```

O–ring failures by PSI level
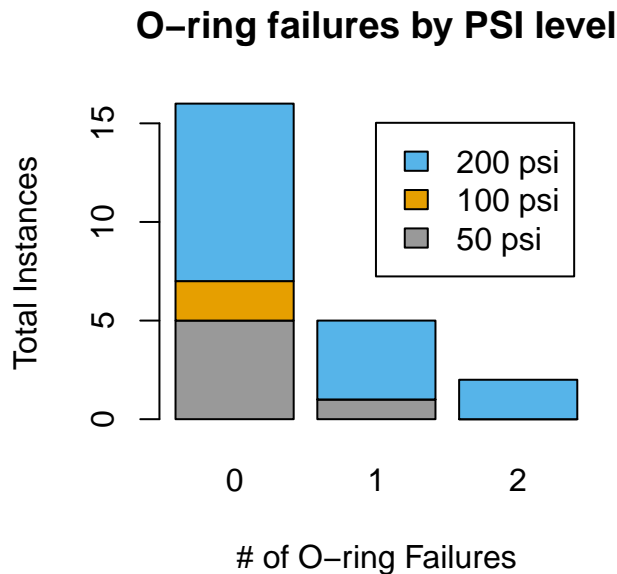
O-ring

- Out of 23 total observations, there were 7 failed launches (30.4%) in total, resulting from o-ring failure(s).
- 5 of the 7 failed launches (71.4%) were caused by 1 o-ring failure.
- 2 of the 7 failed launches (28.6%) were caused by 2 o-ring failures.

Correlations

- Optically, we can see that a large proportion of o-ring failures were at lower temperatures ($<=70F$). Statistically, we also see that there is evidence of a moderate negative correlation between Oring and Temperature variables. Using a correlation test, we get a p value of 0.01 (statistically significant), and a moderate negative correlation coefficient of -0.51.

```
temp_ring <- cor.test(df$Temp, df$O.ring, method = "pearson")
cor(df$Temp, df$O.ring, method = "pearson")
```

```
## [1] -0.5111264
```

```
temp_ring$p.value
```

```
## [1] 0.01268176
```

- There is little to no correlation between the Temperature and Pressure variables (0.04 and 0.28 respectively), as well as the Oring and Pressure variables. Both these relationships also generate high p-values (0.86 and 0.19 respectively); hence insignificant correlation.

```
cor(df$Temp, df$Pressure, method = "pearson")
```

```
## [1] 0.03981769
```

```
cor(df$O.ring, df$Pressure, method = "pearson")
```

```
## [1] 0.2846663
```

```
temp_pressure <- cor.test(df$Temp, df$Pressure, method = "pearson")
ring_pressure <- cor.test(df$O.ring, df$Pressure, method = "pearson")
temp_pressure$p.value
```

```
## [1] 0.8568536
```

```
ring_pressure$p.value
```

```
## [1] 0.1880018
```

Part 2 (20 points) (a) The binomial distribution relies on the assumption that each observation per trial is independent of one another. This could be problematic if the failure of one O-ring could impact the integrity of proximate O-rings. That said, out of the 7 flights where at least one O-ring failed, only two of those flights experienced more than one O-ring failing, which suggests this may not be an issue.

(b) We specify two models – a binomial model and a binary model. The binomial model is specified in the form $(O.ring/Number) \sim Temp + Pressure$, where the proportion of failures is expressed as a function of temperature and pressure, with the number of O-rings per flight as the weights. In this specification, we estimated the logit as $logit(\pi) = 2.52 - (0.098) * Temp + (0.008) * Pressure$, where π is the estimated probability of an O ring failure and $logit(\pi)$ is the log odds. However, the Pressure explanatory variable does not pass the 0.05 level of significance (p = 0.269).

The binary model has a transformation on the dependent variable so that 0 signifies no O-rings failed and 1 signifies at least one O-ring failed. The insights are similar to the binomial model, where the coefficient estimates on temperature are negative and statistically significant, whereas the coefficient estimates on pressure are not significant at the 5% level (p = 0.247). This model is estimated as $logit(\pi) = 13.29 - (-0.229) * Temp + (0.01) * Pressure$.

We also attempted other model specifications to include interactions and quadriatic relationships. While the overall residual deviance declined in the full model with interactions and quadriatic terms, none of the individual explanatory variables alone were statistically significant at the 0.05 level (see model below labeled 'model_fit_full').

```
## transformation to allow for binary logistic regression
df$O.ring.bin <- as.integer(df$O.ring > 0)
## model 1: binomial model, where we fit failures and trials, weighted by trials
## includes Temp and Pressure
model.fit <- glm(formula = O.ring/Number ~ Temp + Pressure, data = df,
                 family = binomial(link = logit), weights = Number)
summary(model.fit)
```

```
##
## Call:
## glm(formula = O.ring/Number ~ Temp + Pressure, family = binomial(link = logit),
##     data = df, weights = Number)
```

6

```
## 
## 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 *
## 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
## model 2: binary model, where we fit the probability that any O-rings fail
## includes Temp and Pressure
model.fit.bin <- glm(formula = O.ring.bin ~ Temp + Pressure, data = df,
                     family = binomial(link = logit))
summary(model.fit.bin)
```

```
## 
## Call:
## glm(formula = O.ring.bin ~ Temp + Pressure, family = binomial(link = logit),
##     data = df)
## 
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1993  -0.5778  -0.4247   0.3523   2.1449
## 
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 13.292360   7.663968   1.734   0.0828 .
## Temp        -0.228671   0.109988  -2.079   0.0376 *
## Pressure     0.010400   0.008979   1.158   0.2468
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 28.267  on 22  degrees of freedom
## Residual deviance: 18.782  on 20  degrees of freedom
```

```
## AIC: 24.782
##
## Number of Fisher Scoring iterations: 5
## binomial model with interactions and quadriatic terms
## includes Temp, Pressure, quadratic terms for both variables, and interactions
model.fit.full <- glm(formula = O.ring/Number ~ Temp + I(Temp^2) +
                          Pressure + I(Pressure^2) + Temp:Pressure, data = df,
                      family = binomial(link = logit), weights = Number)
summary(model.fit.full)
```

```
##
## Call:
## glm(formula = O.ring/Number ~ Temp + I(Temp^2) + Pressure + I(Pressure^2) +
##     Temp:Pressure, family = binomial(link = logit), data = df,
##     weights = Number)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.00828  -0.68897  -0.46399  -0.00018   2.32327
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)     5.360e-01  6.527e+03   0.000    1.000
## Temp            3.579e-01  1.119e+00   0.320    0.749
## I(Temp^2)       9.550e-04  6.141e-03   0.156    0.876
## Pressure       -6.373e-01  1.632e+02  -0.004    0.997
## I(Pressure^2)   3.380e-03  6.527e-01   0.005    0.996
## Temp:Pressure  -2.908e-03  3.536e-03  -0.822    0.411
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 24.23  on 22  degrees of freedom
## Residual deviance: 15.10  on 17  degrees of freedom
## AIC: 40.66
##
## Number of Fisher Scoring iterations: 17
```

(c) We used LRT to test a number of models against the null hypothesis of a single predictor for all observations and found that Temp appears to be significant in all linear specifications of the model (in the most basic specification, null model has residual deviance of 24.23 while model with Temp has 18.086, with a P-value of 0.0132), while Pressure does not. As well, including an interaction term for Temp and Pressure does not appear to be significant relative to the full linearly specified model including Temp and Pressure, which suggests that this term should not be included in the model specification.

```
mod.fit.Ho <- glm(formula = O.ring/Number ~ 1, data = df,
                  family = binomial(link = logit), weights = Number)
mod.fit.Ha_Temp <- glm(formula = O.ring/Number ~ Temp, data = df,
```

```
                      family = binomial(link = logit), weights = Number)
mod.fit.Ha_Pressure <- glm(formula = O.ring/Number ~ Pressure, data = df,
                      family = binomial(link = logit), weights = Number)
mod.fit.Ha <- glm(formula = O.ring/Number ~ Temp + Pressure, data = df,
                      family = binomial(link = logit), weights = Number)
mod.fit.Ha_int <- glm(formula = O.ring/Number ~ Temp  + Pressure + Temp:Pressure,
                      family = binomial(link = logit), weights = Number, data = df)

## ANOVA LRT tests
print("FULL MODEL LRT ONE VARIABLE AT A TIME")
```

```
## [1] "FULL MODEL LRT ONE VARIABLE AT A TIME"
```

```
Anova(mod.fit.Ha)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: O.ring/Number
##          LR Chisq Df Pr(>Chisq)
## Temp       5.1838  1     0.0228 *
## Pressure   1.5407  1     0.2145
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
print("FULL MODEL LRT VS NULL MODEL")
```

```
## [1] "FULL MODEL LRT VS NULL MODEL"
```

```
anova(mod.fit.Ho, mod.fit.Ha, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: O.ring/Number ~ 1
## Model 2: O.ring/Number ~ Temp + Pressure
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        22     24.230
## 2        20     16.546  2   7.6847  0.02144 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
print("TEMP ONLY LRT VS NULL MODEL")
```

```
## [1] "TEMP ONLY LRT VS NULL MODEL"
```

```
anova(mod.fit.Ho, mod.fit.Ha_Temp, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: O.ring/Number ~ 1
## Model 2: O.ring/Number ~ Temp
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
```

```
## 1         22    24.230
## 2         21    18.086  1   6.144  0.01319 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
print("PRESSURE ONLY LRT VS NULL MODEL")
```

```
## [1] "PRESSURE ONLY LRT VS NULL MODEL"
```

```
anova(mod.fit.Ho, mod.fit.Ha_Pressure, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: O.ring/Number ~ 1
## Model 2: O.ring/Number ~ Pressure
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        22      24.23
## 2        21      21.73  1   2.5009   0.1138
```

```
print("FULL MODEL LRT VS FULL MODEL WITH INTERACTION TERM")
```

```
## [1] "FULL MODEL LRT VS FULL MODEL WITH INTERACTION TERM"
```

```
anova(mod.fit.Ha, mod.fit.Ha_int, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: O.ring/Number ~ Temp + Pressure
## Model 2: O.ring/Number ~ Temp + Pressure + Temp:Pressure
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        20      16.546
## 2        19      16.257  1  0.28875   0.591
```

(d) We believe Pressure was removed from the model because a variety of tests for significance and model specification indicate Pressure is not a significant explanatory variable of likelihood of O-ring failure. These include:

- Wald test: the Z-statistic for the Pressure parameter estimate is 1.105 (p = 0.269), which is too small to reject the null hypothesis that the estimator for Pressure is zero
- LRT tests: no matter how Ha and Ho are specified (e.g., whether Ho includes or other explanatory variables as significant estimates), the LR statistic for the Ha with Pressure is never significant

That said, the Z-statistic and corresponding p-value from the logistic regression is not insignificant enough to completely rule out the possibility that Pressure is a meaningful explanatory variable, particularly considering the small sample size on which this regression is estimated. Furthermore, the authors of the paper mention an analysis where bootstrapped confidence intervals were developed at two different pressure points, and these confidence intervals did not completely overlap. This suggests to us that Pressure does not make much of an impact on the confidence interval estimates at the extreme ends of Pressure within the data. In the end, the safest course to take is to drop Pressure as an explanatory variable without additional evidence that its effect is significant on probability of O-ring failure.

Part 3 (35 points)

(a) Consider the simplified model

$$logit(\pi) = \beta_0 + \beta_1 Temp$$

where π is the probability of an O-ring failure. We estimate the logistic regression model using the generalized linear model. In particular, we use the logit link to describe the linear relationship between the log odds of the estimated probability to the temperature. We find that the p-value for this model is 0.014 (< 0.05) and it provides strong evidence to reject the null hypothesis (null: no effect of temperature).

```
log_model_binomial <- glm(formula=O.ring/Number ~ Temp, weights=Number,
                          family=binomial(link=logit), data=df)
summary(log_model_binomial)
```

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

We obtain the following model that describes the probability of an O ring failure π:

$$\pi = \frac{\exp(5.085 - 0.1156 * Temp)}{1 + \exp(5.085 - 0.1156 * Temp)}$$

(b) We can now construct plots for the estimated probability of an O ring failure. In the space rocket, since there are six O rings, we estimate the expected number of O ring failures in the launch to simply be $6 * \pi$. This is possible due to the assumption of independence of failures across O rings within the same spacecraft.

```
n_trials <- 6
w <- aggregate (formula = O.ring ~ Temp, data = df, FUN=sum)

par(mfrow=c(1,2))     # set the plotting area into a 1*2 array
plot (x = w$Temp, y = w$O.ring/ n_trials, xlab = 'Temperature (deg F)',
      ylab = " Estimated probability of failure",
      xlim=c(31,81), ylim=c(0,1), panel.first=grid (col = "gray", lty = "dotted"))
curve(expr=predict(object=log_model_binomial,
                   newdata = data.frame(Temp = x), type="response"),
      col='red', add=TRUE, xlim=c(31, 81),
      ylab = " Estimated probability of failure", xlab='Temperature (deg F)')
legend(x="topright",
       legend=c("Observed proportion of failures", "Estimated probability of failure"),
       col=c("black","red"), lty= c("blank", "solid"), pch=c("o",""))

plot(x = df$Temp, y = df$O.ring, xlab = 'Temperature (deg F)',
     ylab = "Expected Number of Observed Failures",
     xlim=c(31,81), ylim=c(0,6), panel.first=grid (col = "gray", lty = "dotted"))
curve(expr=predict(object=log_model_binomial,
                   newdata = data.frame(Temp = x), type="response")*n_trials,
      col='red', add=TRUE, xlim=c(31, 81),
      ylab = "Expected Number of Failures", xlab='Temperature (deg F)')
legend(x="topright", legend=c("Observed failures", "Estimated failures"),
       col=c("black","red"), lty= c("blank", "solid"), pch=c("o",""))
```
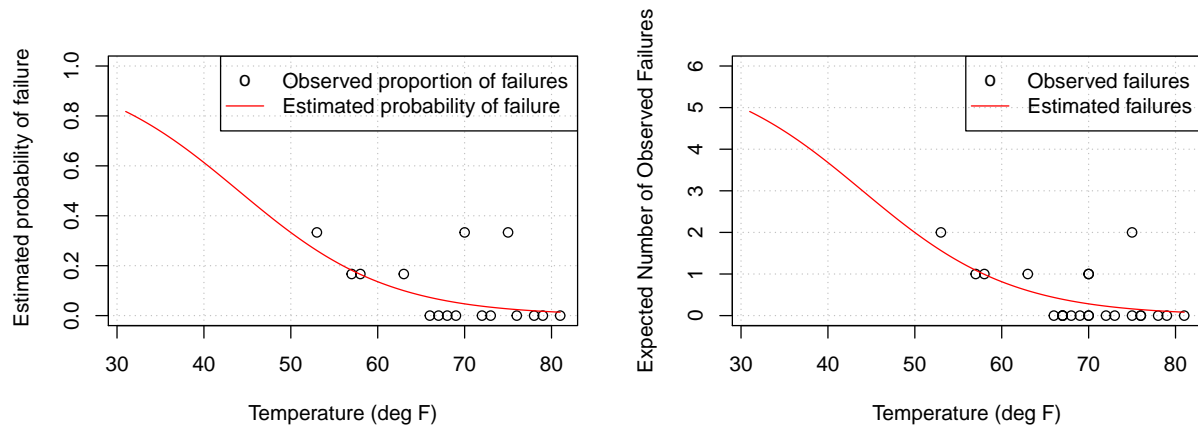

(c) We can further evaluate the confidence intervals on the probability estimates at each temperature of interest. We use the 95% Wald confidence intervals as follows:

- Compute the lower and upper confidence intervals for the logit using the estimated logistic regression coefficients.

$$logit = \hat{\beta}_0 + \hat{\beta}_1 * Temp$$

$$logit.CI.lower(Temp) = (\hat{\beta}_0 + \hat{\beta}_1 * Temp) - Z_{1-\alpha/2}\sqrt{Var(\hat{\beta}_0 + \hat{\beta}_1 * Temp)}$$

$$logit.CI.upper(Temp) = (\hat{\beta}_0 + \hat{\beta}_1 * Temp) + Z_{1-\alpha/2}\sqrt{Var(\hat{\beta}_0 + \hat{\beta}_1 * Temp)}$$

- Use the lower and upper estimates of the logit to derive confidence intervals for the probability of failure

$$pi.CI(lower) = \frac{\exp(logit.CI.lower)}{(1 + \exp(logit.CI.lower))}$$

$$pi.CI(upper) = \frac{\exp(logit.CI.upper)}{(1 + \exp(logit.CI.upper))}$$

We observe that the confidence bounds diverge at lower temperatures. This is observed since not enough observed data is available at lower temperatures.

```
# Function that returns confidence bounds for any model and significance alpha
ci.pi <- function(newdata, mod.fit.obj, alpha) {
  linear.pred <- predict(object = mod.fit.obj,
                         newdata =newdata, type = "link", se = TRUE)
  CI.lin.pred.lower <- linear.pred$fit -
    qnorm (p =  1 - alpha /2) * linear.pred$se
  CI.lin.pred.upper <- linear.pred$fit +
    qnorm (p = 1 - alpha /2) * linear.pred$se
  CI.pi.lower <- exp (CI.lin.pred.lower)/ (1 +  exp (CI.lin.pred.lower))
  CI.pi.upper <- exp (CI.lin.pred.upper)/ (1 +  exp (CI.lin.pred.upper))
  list (lower = CI.pi.lower, upper = CI.pi.upper)
}
plot(x = w$Temp, y = w$O.ring/n_trials, xlab = 'Temperature (deg F)',
     xlim=c(31,81), ylim=c(0,1), ylab = "Estimated Probability of Failure",
     panel.first=grid(col = "gray", lty = "dotted"))
curve(expr=predict(object=log_model_binomial,
                   newdata = data.frame(Temp = x), type="response"),
      col='red', add=TRUE, xlim=c(31, 81), ylim=c(0,1),
      ylab = "Estimated Probability of Failure", xlab='Temperature (deg F)')
curve (expr = ci.pi(newdata = data.frame(Temp = x),
                    mod.fit.obj = log_model_binomial, alpha = 0.05)$lower,
       col = "blue", lty="dotdash", add = TRUE, xlim = c(31, 81), ylim=c(0,1),
       ylab = 'Estimated Probability of Failure',xlab='Temperature (deg F)')
curve (expr = ci.pi (newdata = data.frame(Temp = x),
                     mod.fit.obj = log_model_binomial, alpha = 0.05)$upper,
       col = "blue", lty= "dotdash", add = TRUE,
       xlim = c(31,81), ylim=c(0,1), xlab='Temperature (deg F)')
legend(x="topright",
       legend=c("Logistic regression model",
                "Wald 95%  individual C. I.", "Observed Failure Proportions"),
       pch=c("","","o"),
       lty= c("solid", "dotdash", "blank"), col = c("red", "blue", "black"))
```
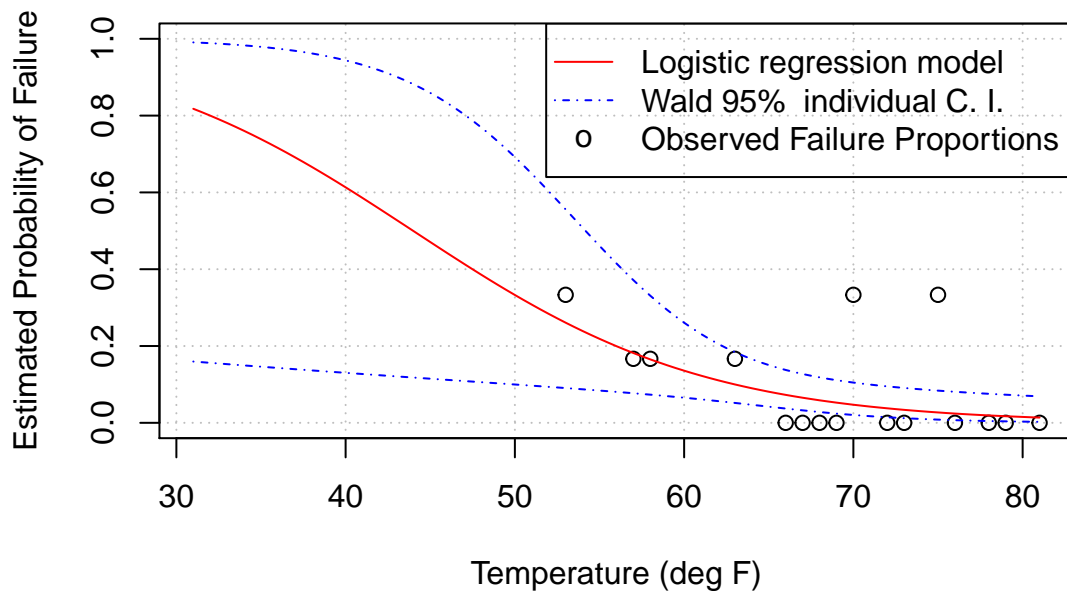
The temperature was 31° at launch for the Challenger in 1986. We now estimate the probability of an O-ring failure at this temperature based on the models developed above. We compute the Wald confidence intervals and assume that the failure of each O ring in the space rocket is independent to the other. To apply the model to the Challenger takeoff at 31F, we assume the following:

- The O rings used in the Challenger are similar to the ones used before (material, size)
- The thermal stresses on the O rings are similar to those on preceding launches
- Each O ring success/failure is independent of the outcomes of other O rings in the Challenger.

The estimated probability of an O ring failure at 31F is 0.8178. It is estimated that approximately 5 ($0.8178 * 6 = 4.9$) O rings could fail. The 95% Wald confidence intervals for the 31F launch is ($0.159 - 0.9906$). In other words, the launch could have any where between 1 to 6 O rings fails.

```
temperature = 31
prob<- predict(object=log_model_binomial,
               newdata=data.frame(Temp=c(temperature)), type='response')
print(paste0('Estimated probability of an O ring failure at 31F: ',
             round(prob[1][1],3)))
```

```
## [1] "Estimated probability of an O ring failure at 31F: 0.818"
```

```
print(paste0('Estimated number of O ring fails: ',
             round(prob[1][1]*6,0)))
```

```
## [1] "Estimated number of O ring fails: 5"
```

```
ci <- ci.pi(newdata=data.frame(Temp=c(temperature)),
            mod.fit.obj=log_model_binomial, alpha=0.05)
print(paste0('Estimated 95% confidence bound for probability of an O ring failure at 31F:'))
```

```
## [1] "Estimated 95% confidence bound for probability of an O ring failure at 31F:"
print(paste0('(', round(ci$lower,3), ',', round(ci$upper,3), ')'))
```

```
## [1] "(0.16,0.991)"
print(paste0('Estimated 95% confidence bound for number of O ring failures'))
```

```
## [1] "Estimated 95% confidence bound for number of O ring failures"
print(paste0('in the spacecraft at 31F: (',
             round(ci$lower*6,0), ',', round(ci$upper*6,0), ')'))
```

```
## [1] "in the spacecraft at 31F: (1,6)"
temperature = 72
prob<- predict(object=log_model_binomial,
               newdata=data.frame(Temp=c(temperature)), type='response')
print(paste0('Estimated probability of an O ring failure at 72F: ',
             round(prob[1][1],3)))
```

```
## [1] "Estimated probability of an O ring failure at 72F: 0.038"
```

(e) **Parametric Bootstrap**: The confidence intervals could also be computed using bootstrap sampling method. In Section 3.2 of the paper, the authors describe an alternative way of generating synthetic data where they use the estimated model to generate new binomial responses. For this to work, we need to assume that the model we first estimated is close to being a true model. We sample the temperatures from the test data and synthetically generate a binomial response for each temperature using the estimated model from above. We create 1000 datasets of 23 sampled temperatures each, and each dataset with the candidate model. We have now generated distributions for the estimates β_0 and β_1. We use the distributions $(5\% - 95\%)$ get the estimates for 90% CI probability of failure at 31F to be $(0.589 - 0.903)$ and $(0.008 - 0.099)$ at 72F. This method seems to generate a tighter confidence bound for the lower temperatures than other bootstrap methods (see below).

```
suppressWarnings({
beta0 = log_model_binomial$coefficients[1]
beta1 = log_model_binomial$coefficients[2]
num_datasets = 1000

run_monte_carlo <- function(num_samples_per_dataset, xlimit, params){
  temperature <- sample(xlimit, num_samples_per_dataset, replace=T)
  temperature = rep(temperature, times=6)
  beta0 = params[1]
  beta1 = params[2]
  pi <- exp (beta0 + beta1 * temperature)/ (1 + exp (beta0 + beta1 * temperature))
  y <- rbinom(n = length(temperature), size = 1, prob = pi)
  df_sim <- data.frame(y=y, temperature=temperature)
  mod.fit <- glm(formula = y~ 0 + temperature + offset(rep(beta0, length(temperature))),
                 data=df_sim, family=binomial(link=logit))
  beta.hat1 <- mod.fit$coefficients[1]
  mod.fit.1<-glm(formula=y~offset(beta1*temperature),data=df_sim,family=binomial(link=logit))
```

15

```
    beta.hat0 <- mod.fit.1$coefficients[1]
    list(beta0=beta.hat0, beta1=beta.hat1, model=mod.fit)
}
num_samples_per_dataset = 23
params = c(beta0, beta1)
beta0_runs <- c()
beta1_runs <- c()
for (i in 1:num_datasets) {
  out <- run_monte_carlo(num_samples_per_dataset, df$Temp, params)
  if (out$model$converged == TRUE) {
    beta0_runs <- c(beta0_runs,out$beta0)
    beta1_runs <- c(beta1_runs,out$beta1)
  }
}
beta_0_quantiles = quantile(beta0_runs, c(0.05, 0.95))
beta_1_quantiles = quantile(beta1_runs, c(0.05, 0.95))
tt <- beta_0_quantiles + beta_1_quantiles*31
pi_31 = exp(tt)/ (1+exp(tt))
tt <- beta_0_quantiles + beta_1_quantiles*72
pi_72 = exp(tt)/ (1+exp(tt))

print("The confidence bounds for estimated probability of an O ring failure at 31F:")
print(pi_31)
print("The confidence bounds for estimated probability of an O ring failure at 72F:")
print( pi_72)
})
```

```
## [1] "The confidence bounds for estimated probability of an O ring failure at 31F:"
##         5%        95%
## 0.5978389 0.9044777
## [1] "The confidence bounds for estimated probability of an O ring failure at 72F:"
##           5%          95%
## 0.008048377 0.101588552
```

Basic Bootstrap: To avoid the true model assumption, we prefer to use the basic bootstrapping method and sample the experimental dataset over and over again (1000 repeats with replacement) in batches of 23. This simulates several spacecraft launches at the operational temperatures but examined in batches of 23. We then fit new models to each of the bootstrapped dataset (n=23) and compute the estimated probability of O ring failures at 31F and 72F for each of the datasets. This results in a distribution of estimated probability of failures at the two chosen temperatures. We finally pick the 5% and 95% quantiles as our confidence bounds. The 90% confidence bounds for the estimated probability of O ring failure using the bootstrapped model at 31F is $(0.227 - 0.993)$, as opposed to a 95% Wald Confidence interval of $(0.159 - 0.9906)$. Using this model, it is estimated that at 31F there can be anywhere between 1 to 6 O ring failures. This is a similar conclusion to that of the 95% Wald confidence interval estimate. On the other hand, at 72F we estimate 0 to 1 fails (estimated prob: $0.006 - 0.078$). The odds of failure of an O ring would have reduced by 99%

had the launch been postponed to 72F.

$$\frac{odds|_{72F}}{odds|_{31F}} = \frac{\exp(5.085 - 0.1156 * 72)}{\exp(5.085 - 0.1156 * 31)} \approx 0.008$$

```r
suppressWarnings({
set.seed(1)
bootstrap_results<- c()
test_temperature <- data.frame(Temp = c(31,72))
num_samples_per_dataset <- 23
num_datasets<-1000
for (i in 1:num_datasets) {
  sample_rows <- sample(row.names(df), num_samples_per_dataset,
                        replace = TRUE)
  bootstrap_test_data <- df[sample_rows , ]
  # generate model for bootstrapped sample
  model <- glm(O.ring/Number ~ Temp, data = bootstrap_test_data,
              weights=Number, family = binomial)
  beta1.hat <- model$coefficients[2]
  # estimate probabilities at test temperatures
  prediction <- predict(model, test_temperature)
  prediction <- exp(prediction) /( 1 + exp(prediction))
  bootstrap_results<- append(bootstrap_results, prediction)
}

bootstrap_results <- matrix(bootstrap_results, nrow = num_datasets, byrow = TRUE)
print("The confidence bounds for estimated probability of an O ring failure at 31F:")
print(c(quantile(bootstrap_results[,1], 0.05),quantile(bootstrap_results[,1], 0.95)))
print("The confidence bounds for estimated probability of an O ring failure at 72F:")
print( c(quantile(bootstrap_results[,2], 0.05),quantile(bootstrap_results[,2], 0.95)))
 })
```

```
## [1] "The confidence bounds for estimated probability of an O ring failure at 31F:"
##        5%       95%
## 0.2267247 0.9937941
## [1] "The confidence bounds for estimated probability of an O ring failure at 72F:"
##         5%        95%
## 0.00566118 0.07877557
```

(f) We use the quadratic term in temperature to fit a new model. We find that the change in residual deviance is 0.495 with a degree of freedom = 1. The X^2 distribution for a 95% significance with $df = 1$ requires the change to be more than 3.841. We therefore conclude that the quadratic term is insignificant.

```r
#Quadratic model
mod.quad.fit <- glm(formula=O.ring/Number ~ Temp + I(Temp^2),
                   family=binomial(link=logit), weights=Number, data=df)
anova(log_model_binomial, mod.quad.fit, test='Chisq')
```

```
## Analysis of Deviance Table
```

```
##
## Model 1: O.ring/Number ~ Temp
## Model 2: O.ring/Number ~ 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
```

Part 4 (10 points)

We will opt to use a binary logistic regression as opposed to a linear regression model in this case for 3 main reasons.

First, we violate a number of the 6 assumptions required for a linear regression model to be valid (details below). Second, a linear model does not capture incremental differences the same way a logistic regression model does. For example, we would expect there to be a significantly higher failure rate if our `temperature` variable dropped 10 degrees from 40 to 30 degrees F vs. dropping 10 degrees from 80 to 70 degrees F. A linear model weights any 10 degree drop equally, whereas the logistic model would capture this nonlinear relationship at lower temperatures. Lastly, the outcome variable (predicted values) aren't bound between 0 and 1 in a linear regression model (whereas it is bounded between 0 and 1 in a binary logistic regression model).

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

```
##
## Call:
## lm(formula = O.ring.bin ~ Temp, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.43762 -0.30679 -0.06381  0.17452  0.89881
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.90476    0.84208   3.450  0.00240 **
## Temp        -0.03738    0.01205  -3.103  0.00538 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3987 on 21 degrees of freedom
## Multiple R-squared:  0.3144, Adjusted R-squared:  0.2818
## F-statistic:  9.63 on 1 and 21 DF,  p-value: 0.005383
```

```r
par(mfrow=c(2,2))
plot(lm.model)
```

18
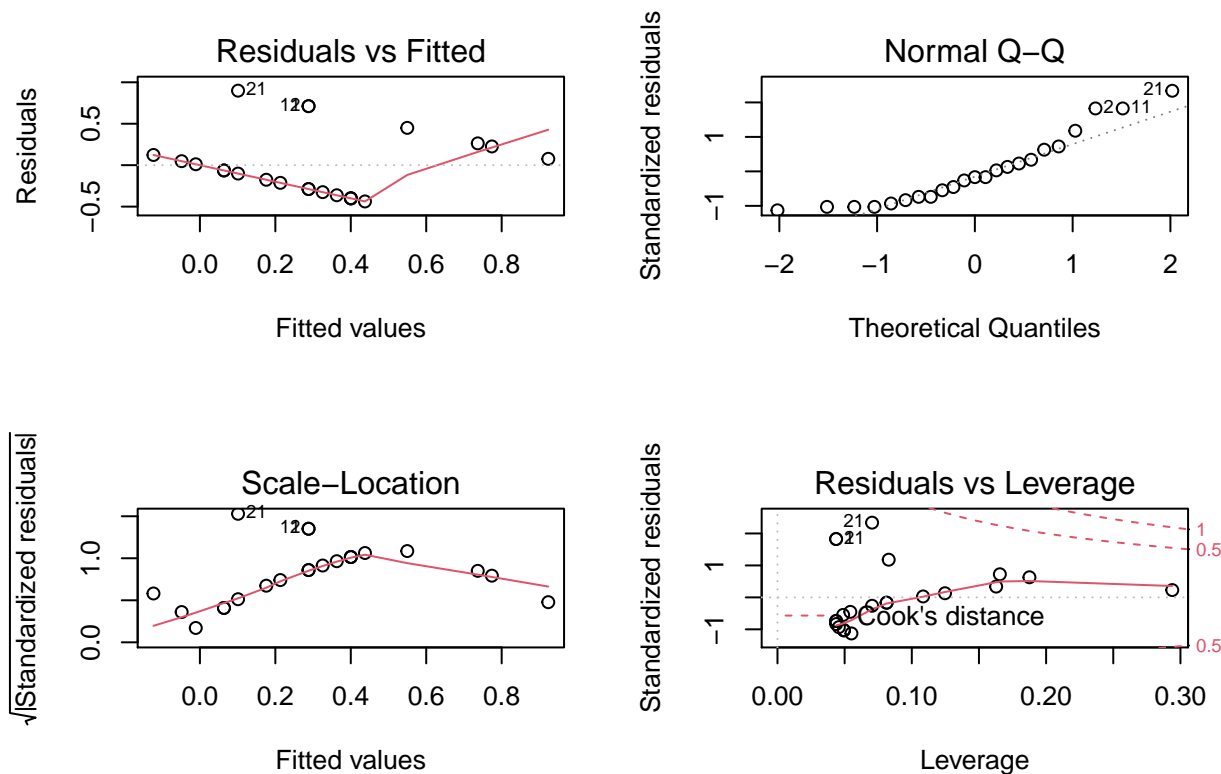
At least 3 out of 6 requirements for the linear regression model are violated (below). The 6 assumptions for the linear regression model are:

1. Linear in Parameters - Looking at the first "Residuals vs. Fitted" plot, we see that this is violated. Ideally we would like this line to be a flat, horizontal line, but we see that most of the points on that line create a negative sloping line, and the rest are not scattered evenly above and below the hypothetical horizontal line at 0.

2. IID - The iid requirement is satisfied for the following reasons:

- Per the paper, the authors assumed each rocket launch is independent.
- Out of the flights where at least one o-ring failed, only two flights had more than one o-ring fail (less than half). If one o-ring failure foreshadowed another o-ring failure (o-ring failures were not independent), we would have seen a higher proportion of failures with 1+ o-ring failures. This is not the case so we do see independence here.
- When we look at the "Residuals vs. Leverage" plot, we see a flat horizontal line for the most part (preferred).

3. No perfect multicollinearity - Satisfied since we only have 1 explanatory variable.

4. Zero conditional mean of residuals - Violated based on the "Residuals vs Fitted" graph above. Ideally we would like this line to be a flat, horizontal line, but we see that most of the points on that line create a negative sloping line, and the rest are not scattered evenly above and below the hypothetical horizontal line at 0.

5. Homoskedasity of errors - Violated based on the "Scale vs Location" graph above. Ideally,

19

we would like to see a horizontal band of points. However, we have what looks like an upside down parabola, something that signals we may have heteroskedacity.

6. Errors are normally distributed - Based on the "Normal Q-Q" graph above, we have a plot that could be argued either way. We simply do not have enough data points to conclude if we satisfy or violate this assumption. Most of the residuals fall along the linear, diagonal line. However, there is some deviation at the extremes towards the beginning of the line, and especially towards the end where there are not that many samples.

Part 5 (10 points)

Part 1: We performed EDA on a very small but clean dataset (23 observations each for 3 main variables: Temperature, Pressure, O-ring). A critical finding was that there is a significant negative relationship between the Temperature and O-ring variables. Analysis performed on the variables, however, did not yield much significance - mostly inconclusive or insignificant relationships.

Part 2: In its binomial form, the model suggests that a 1 degree increase of temperature is associated with a 0.098 decrease in the log odds of the failure of any individual ring. Alternatively, a 1 degree increase in temperature is associated with a 0.22 decrease in the log odds of at least one ring failing. In terms of probabilities, a 1 degree increase in temperature is associated with a 0.90 impact on the odds of any O-ring failure. We cannot conclude that Pressure has a meaningful impact on likelihood of O-ring failure due to the high P-values (>0.20) associated with both the binary and binomial model estimating the impact of Pressure on likelihood of failure.

Part 3: We finally chose the binomial model $O.ring/Number \sim Temperature$ based on the analyses performed. We find that the estimated probability of an O ring failure for a 31F launch is 0.8178 and the probability of at least one O ring failure during launch to be at 99.99%

$$P(At \ least \ 1 \ failure) = 1 - P(No \ failure) = 1 - (1 - 0.8178)^6 = 99.99\%$$

If the launch was delayed to when temperature was 72F, we find that the mean estimate for probability of at least one O ring failure during launch reduces to 20.74%

$$P(At \ least \ 1 \ failure) = 1 - P(No \ failure) = 1 - (1 - 0.038)^6 = 20.74\%$$

Note that failure of an O ring does not imply a catastrophic failure of the spacecraft. Finally, the estimated odds of an O ring failure are 114 times as large for launches at 31F than those at 72F.

$$\frac{odds|_{31F}}{odds|_{72F}} = \frac{\exp(5.085 - 0.1156 * 31)}{\exp(5.085 - 0.1156 * 72)} \approx 114$$

We therefore conclude that the best choice for the team would have been to postpone the launch to warmer temperatures since the possibility of O ring failures was significantly high at 31F.

Part 4: We concluded that the logistic regression model is more suitable than the linear model for this exercise. 3 reasons for this conclusion: 1) we violate 3 of the 6 assumptions required for use of the linear model, 2) the linear model does not capture incremental differences the same way a logistic regression model does, and 3) the outcome variable isn't bound between 0 and 1 (hence, produces erroneous predictions).