Statistical Methods for Discrete Response, Time Series, and Panel Data (W271): Lab 1

W271 Legg Yeung, Phat Doan, Ted Pham

June 14, 2017

```
suppressMessages(library(car))
suppressMessages(library(dplyr))
suppressMessages(library(Hmisc))
suppressMessages(library(stargazer))
suppressMessages(library(ggplot2))
suppressMessages(library(moments))
suppressMessages(library(reshape2))
suppressMessages(library(lmtest))
suppressMessages(library(sandwich))
suppressMessages(library(mcprofile))
```

Investigation of the 1989 Space Shuttel Challenger Accident

The failure of an O-ring on the space shuttle Challenger's booster rockets led to its destruction in 1986. Using data on previous space shuttle launches, Dalal et al. (1989), our group has been task to examine the probability of an O-ring failure as a function of temperature at launch and combustion pressure. We use data from Dalal et al.'s paper for this analysis.

2. Exploratory Data Analysis (EDA)

```
df = read.csv("challenger.csv",sep = ",")
```

2.1 Summary of Descriptive Statistical Analysis

```
suppressWarnings(describe(df))
```

```
## df
##
##  5  Variables      23  Observations
## ---------------------------------------------------------------------------
## Flight
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##       23        0       23        1       12        8      2.1      3.2
##      .25      .50      .75      .90      .95
##      6.5     12.0     17.5     20.8     21.9
##
## lowest :  1  2  3  4  5, highest: 19 20 21 22 23
## ---------------------------------------------------------------------------
## 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
## -------------------------------------------------------------------------------
## 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
## -------------------------------------------------------------------------------
## 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
## -------------------------------------------------------------------------------
## Number
##          n  missing distinct      Info      Mean       Gmd
##         23        0        1         0         6         0
##
## Value        6
## Frequency   23
## Proportion   1
## -------------------------------------------------------------------------------
```

- There are 23 observations, each corresponding to a launching mission of the Challenger prior to its accident on January 28th 1986. In total 5 variables are available:

  - FLight : flight number, each corresponding to a launching mission
  - Temp : atmospheric temperature at launch time
  - Pressure : leak-check pressure held at the solid rocket motors
  - O-ring : number of filed-joint primary O-rings that experienced failure in form of erosion or blowby
  - Number : total number of primary O-rings at all six field-joints. Value is 6 for all records because design of the challenger solid rocket motor remained the same

- There are no missing values for all records and variables.

- We are interested in assessing risk of O-ring failures based on launching conditions with regression models.

  - The response variable can be constructed from "O-ring" and "Number" to indicate proportion of O-ring failure of all O-rings at each launch, or constructed from "O-ring" to a binary variable to indicate any failures(1) or no failures(0).

  - The explanatory variable can be "Temp", "Pressure", higher order or interaction terms of them.

2

It would be useful to have varibles of the material property of each O-ring because prior pressure testing on the O-rings could have deformed certain O-rings beyond its elastic range thus facilitate erosion(failure).

2.2 Univariate Analysis

2.2.1 Response Variable: O-ring

- "O-ring" has 3 distinct values of [0,1,2]. A larger proportion – 16 out of the 23 launches experienced 0 failures, 5 experienced 1 failure and 2 experienced 2 failures. In total, 9 out of 138 O-rings (6.5%) experienced failures. The distribution has a mean of 0.39, standard deviation of 0.66, and positively skewed with skewness of 1.40. No outliers are apparent. Our regression residuals can be expected to mirror this pattern.

- Since "Number" is constant across all records, a binomial response varible simply requires a scale adjustment to the x-axis of this distribution plot. This distribution should look the same.

```
describe(df$O.ring)
```

```
## df$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
```

```
sd(df$O.ring)
```

```
## [1] 0.6563764
```

```
moments::skewness(df$O.ring)
```

```
## [1] 1.396202
```

- Also, note that the distribution from observations resemble a binomial simulation with n = 6 and probability of failure equals to 0.065. The simulation is plotted below. This means that the underlying distribution of our dependent variable is reasonable enough for binomial logistic regression to be applied.

```
hist( x = df$O.ring, main = "Distribution of O-ring failures overlap binomial distribution",
      xlab = "Number of O-ring failures", ylab = "Frequency",
      ylim = c(0,1),breaks = 0:5 - 0.5, xaxt = "n", yaxt = "n",
      col = rgb(1, 0, 0, .5), freq = F)

x = rbinom(n = 1000, size = 6, prob = 0.065)
hist(x, breaks = 0:5 - 0.5, ylim = c(0,1), xaxt = "n", yaxt = "n",
     col = rgb(0,0,1,.5), freq = F,
     main = "binomial simulation n = 6, pi = 0.065", add = T)

abline(v = mean(df$O.ring), col = "red")
abline(v = median(df$O.ring), col = "green")
axis(1, at = 0:3)
axis(2, at = seq(0,1, by = 0.1))

legend(2.6,1, c("Mean", "Median"), col = c("red", "green"), pch = "|")
legend(2.6,0.7 , cex = 1, title = "legend", legend = c("observations", "simulation"), horiz = F
, fill = c(rgb(1, 0, 0, .5),rgb(0,0,1,.5)))
```

Distribution of O−ring failures overlap binomial distribution
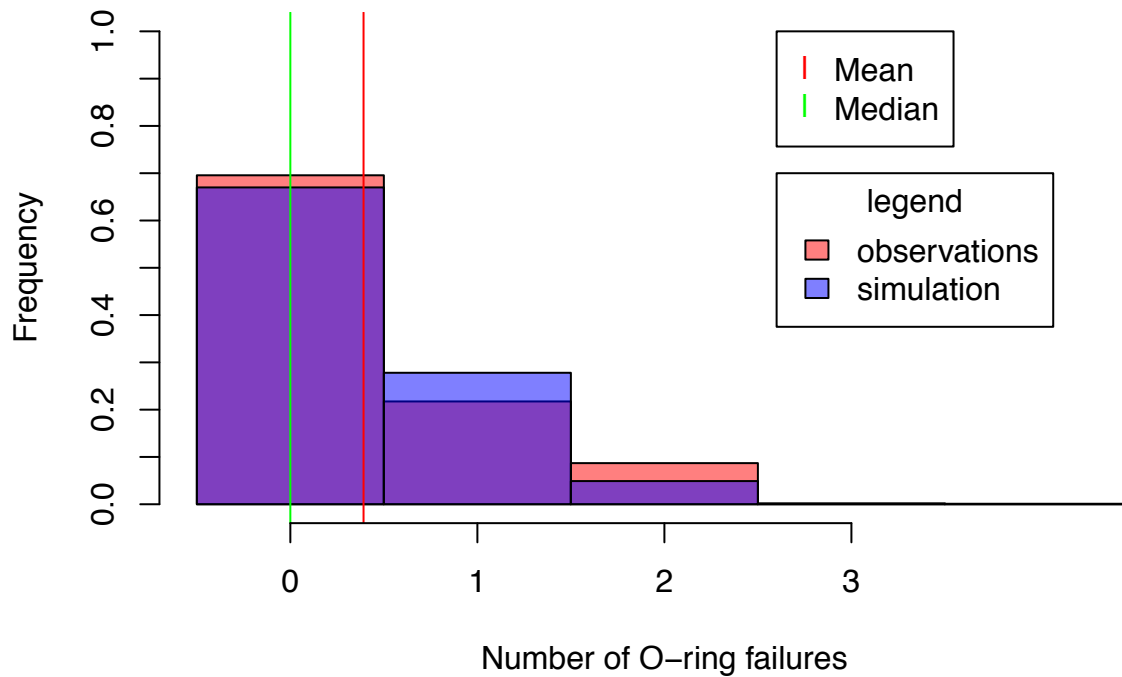
- We may consider transforming this variable into a binary variable. We use 1 to indicate one or more failures at each launch and 0 to indicate no failure at all. This distribution has a mean of 0.304 and a standard deviation of 0.470. This transformation may remove dependence between O-rings failures under the same launch which we will discuss more in section 4.

```
df$O.failed.binary = ifelse(df$O.ring == 0, 0, 1)

describe(df$O.failed.binary)

## df$O.failed.binary
##        n  missing distinct     Info      Sum     Mean      Gmd
##       23        0        2    0.636        7   0.3043   0.4427
sd(df$O.failed.binary)

## [1] 0.470472
hist( x = df$O.failed.binary, main = "Distribution of O-ring failures (Binary)",
      ylab = "Frequency", xlab = "Failure occurence as Binary (0: No | 1: Yes)", ylim = c(0,20),breaks =
      xaxt = "n", yaxt = "n", col = "blue" )
abline(v = mean(df$O.failed.binary), col = "red")
abline(v = median(df$O.failed.binary), col = "green")
axis(1, at = 0:1)
axis(2, at = seq(0,17, by = 1))

legend("topright", c("Mean", "Median"), col = c("red", "green"), pch = "|")
```

Distribution of O–ring failures (Binary)

Failure occurence as Binary (0: No | 1: Yes)
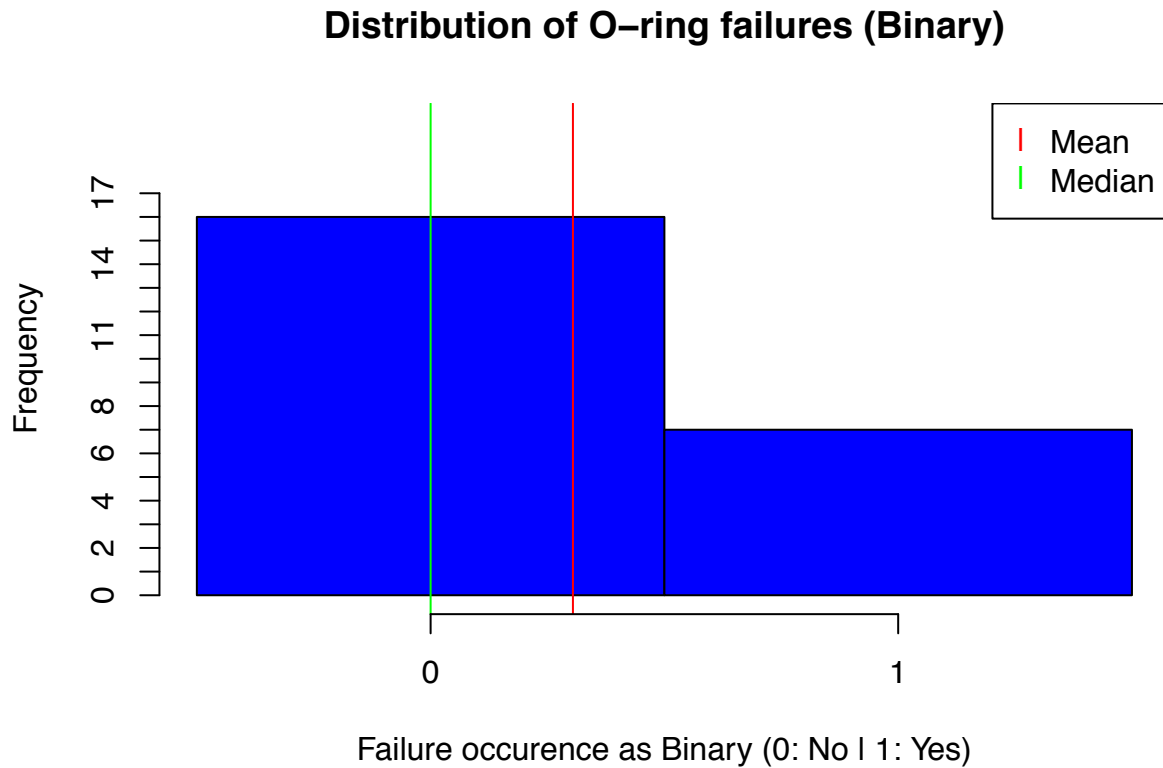
2.2.2 Explanatory Variable: Temperature

- There are 16 distinct values in temperature. It distribution peaks between 65F to 70F with a standard deviation of 7.057. No outliers are apparent because it's a natural process measured at the same site. Negative skew is mild with a skewness of -0.61. We recommend keeping this variable as a continuous variable without log transformation.

```
describe(df$Temp)
```

```
## df$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
```

```
sd(df$Temp)
```

```
## [1] 7.05708
```
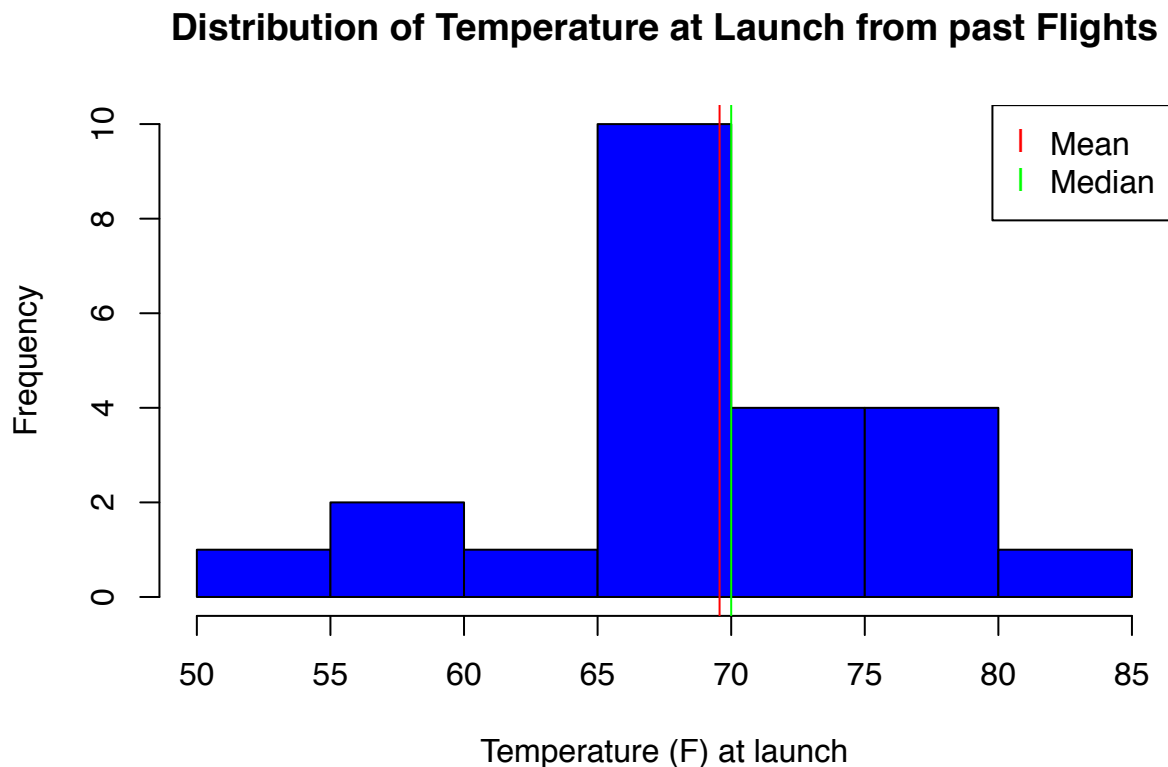
```
moments::skewness(df$Temp)
```

```
## [1] -0.6102761
```

```
hist(df$Temp, main = "Distribution of Temperature at Launch from past Flights"
     ,xlab = "Temperature (F) at launch"
     ,col = "blue", breaks = 8)

abline(v = mean(df$Temp), col = "red")
abline(v = median(df$Temp), col = "green")

legend("topright", c("Mean", "Median"), col = c("red", "green"), pch = "|")
```

Distribution of Temperature at Launch from past Flights

- Higher order polynomial transformation may be considered to compress the negative tail and stretch out the positive tail so to make the distribution less negatively skewed. The squared term of temperature has skewness of -0.364 and the cubic term of temperature has skewness of -0.130.

```
moments::skewness(df$Temp^2)
```
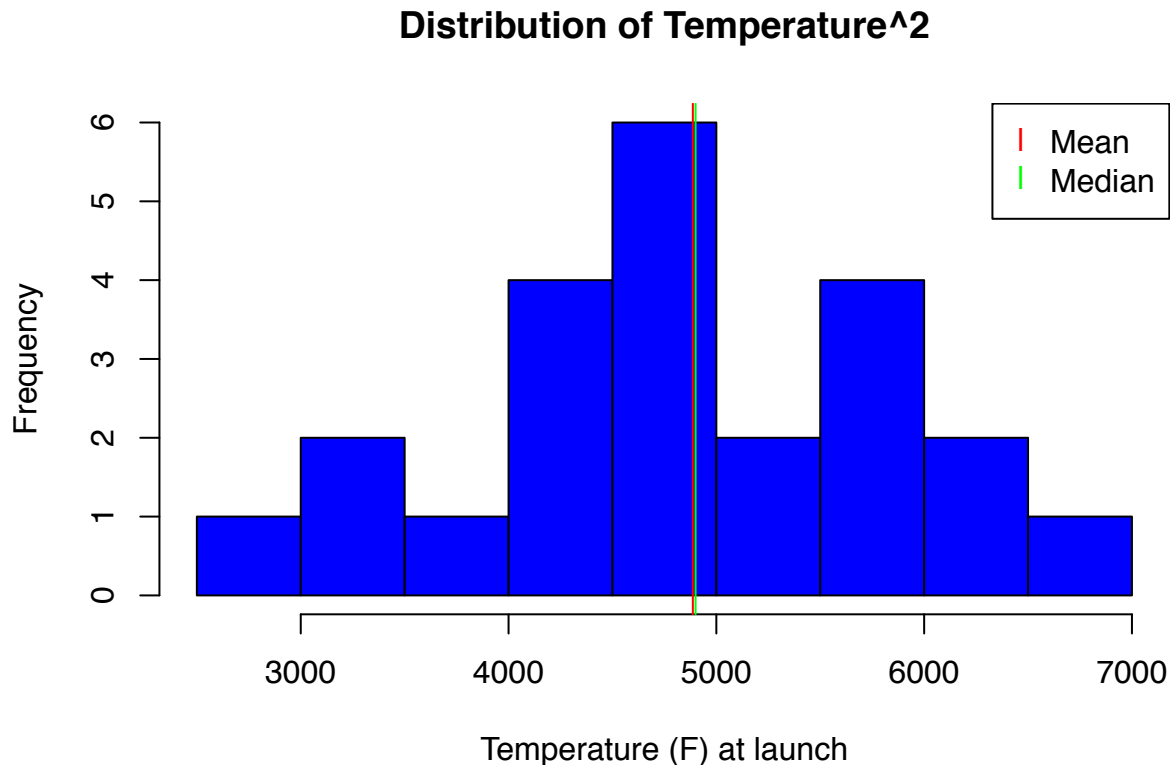
```
## [1] -0.3649353
```

```
moments::skewness(df$Temp^3)
```

```
## [1] -0.1300152
```

```
hist(df$Temp^2, main = "Distribution of Temperature^2"
     ,xlab = "Temperature (F) at launch"
     ,col = "blue", breaks = 8)
```

```
abline(v = mean(df$Temp^2), col = "red")
abline(v = median(df$Temp^2), col = "green")

legend("topright", c("Mean", "Median"), col = c("red", "green"), pch = "|")
```

Distribution of Temperature^2

2.2.3 Explanatory Variable: Pressure

- There are 3 distinct values (50psi, 100psi, 200psi) in pressure. There is no natural continuous variation because leak-check pressure is controled in the solid rocket motors by engineering decisions. Pressure was held at 50psi for the first 6 launches, 100psi for the 7th and 8th launches, and 200 psi for the remaining 15 launches. The distribution has a mean of 152.2psi, standard deviation of 68.221psi and noticable negative skew of -0.738. To assess pressure's effect on O-ring failure, we recommend keeping this variable with its continuous value.

```
describe(df$Pressure)
```

```
## df$Pressure
##        n  missing distinct     Info     Mean      Gmd
##       23        0        3    0.706    152.2    67.59
##
## Value          50   100   200
## Frequency       6     2    15
## Proportion  0.261 0.087 0.652
```
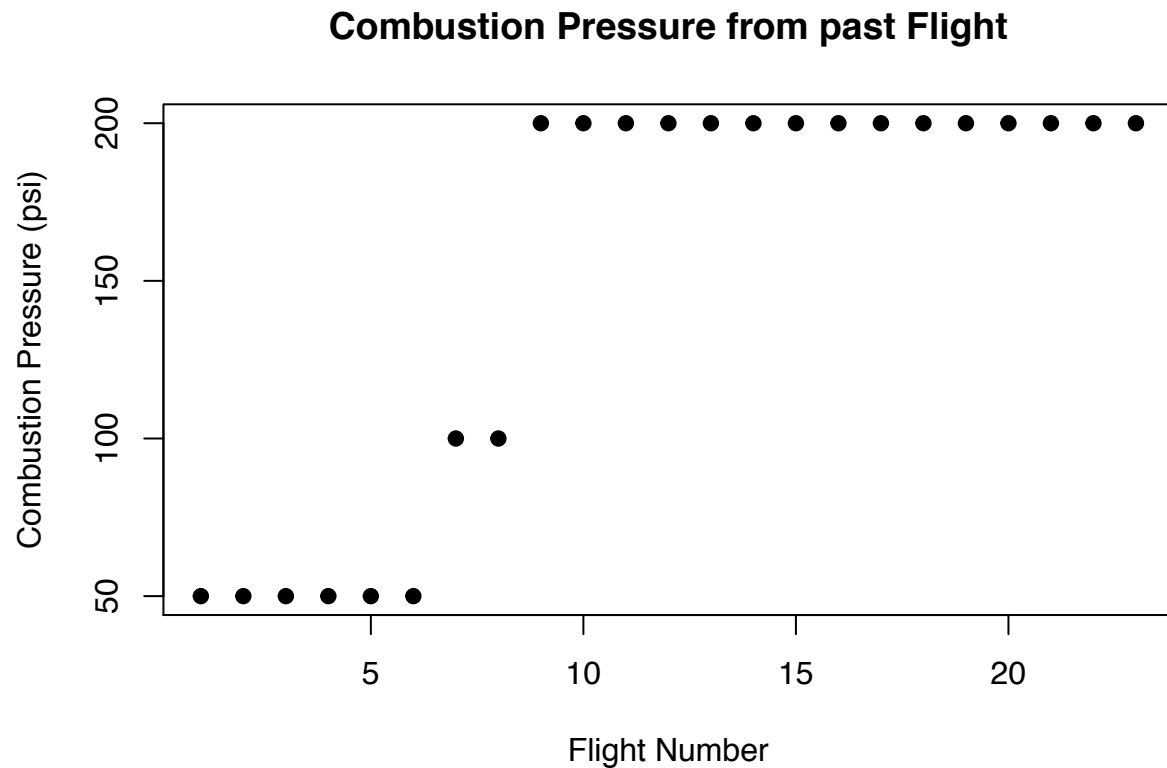
```
sd(df$Pressure)
```

```
## [1] 68.22133
```

7

```
moments::skewness(df$Pressure)
```

```
## [1] -0.738174
```

```
plot(df$Pressure ~ df$Flight, xlab = "Flight Number", ylab = "Combustion Pressure (psi)", main = "Combus
```
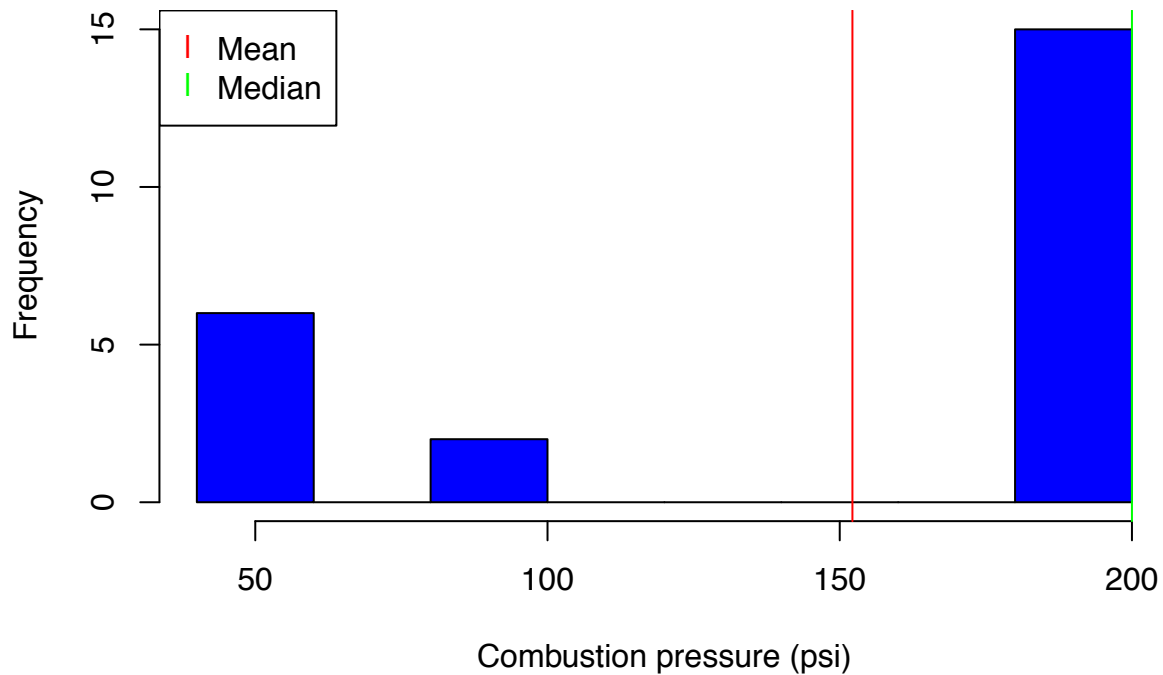
Combustion Pressure from past Flight


```
hist(df$Pressure, main = "Distribution of Combustion Pressure (psi)"
    ,xlab = "Combustion pressure (psi)"
    ,col = "blue")

abline(v = mean(df$Pressure), col = "red")
abline(v = median(df$Pressure), col = "green")

legend("topleft", c("Mean", "Median"), col = c("red", "green"), pch = "|")
```

Distribution of Combustion Pressure (psi)

2.3 Multivariate Analysis

2.3.1 Multivariate Correlation overview

- We notice a couple features from the correlation matrix and scatter matrix which can help guide our analysis:

  - "O.ring" is moderately, negatively correlated with "Temp" (correlation value −0.511). The moderately downward green curve in the scatter matrix confirms.

  - "O.ring" is midly, positively correlated with "Pressure" (correlation value of 0.285). The mildly upward green curve in the scatter matrix confirms.

  - "Temp" is weakly, positively correlated with "Pressure" (correlation value 0.040). The weakly upward green curve in the scatter matrix confirms. This suggests in a regression model, ~~the independent variables have an effect on the dependent variable, without~~ strong collinearity among themselves, which is good. *there will not be*
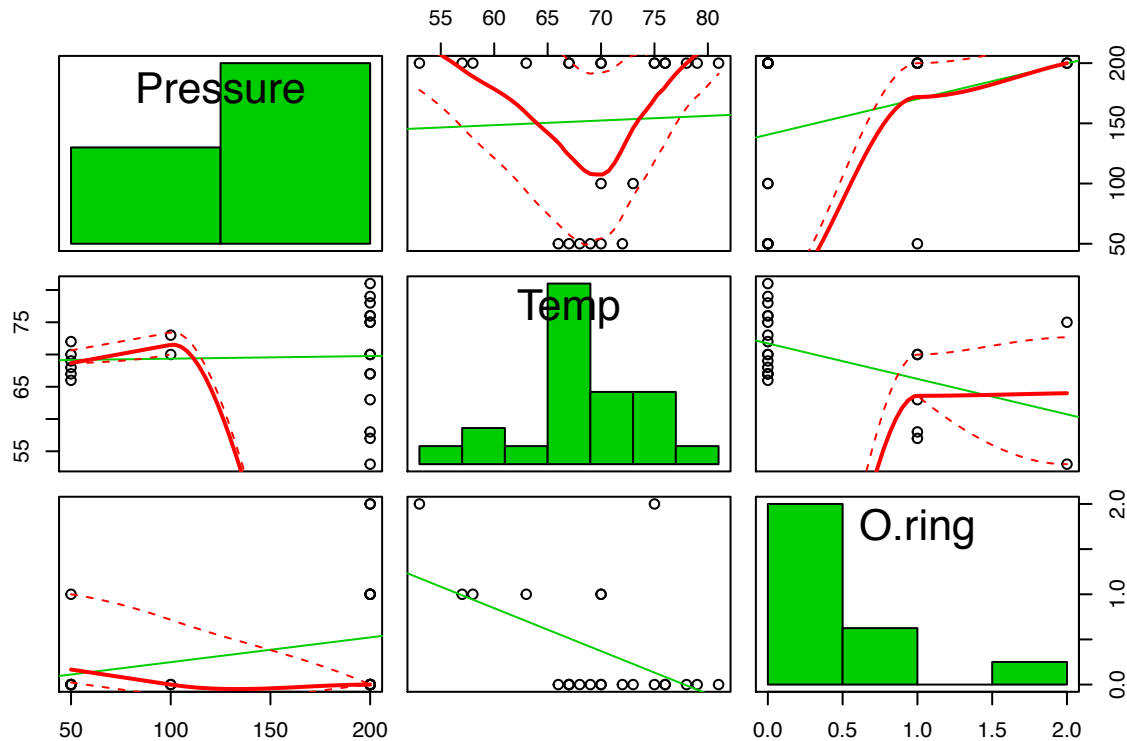
```
vars =  c("Pressure", "Temp", "O.ring")
df1 = df[,vars]

# corraltion matrix
cor(df1)
```

```
##           Pressure       Temp      O.ring
## Pressure 1.00000000  0.03981769  0.2846663
## Temp     0.03981769  1.00000000 -0.5111264
## O.ring   0.28466627 -0.51112639  1.0000000
```

```
# scatterplot
suppressWarnings(scatterplotMatrix(df1, diagonal = "histogram"))
```
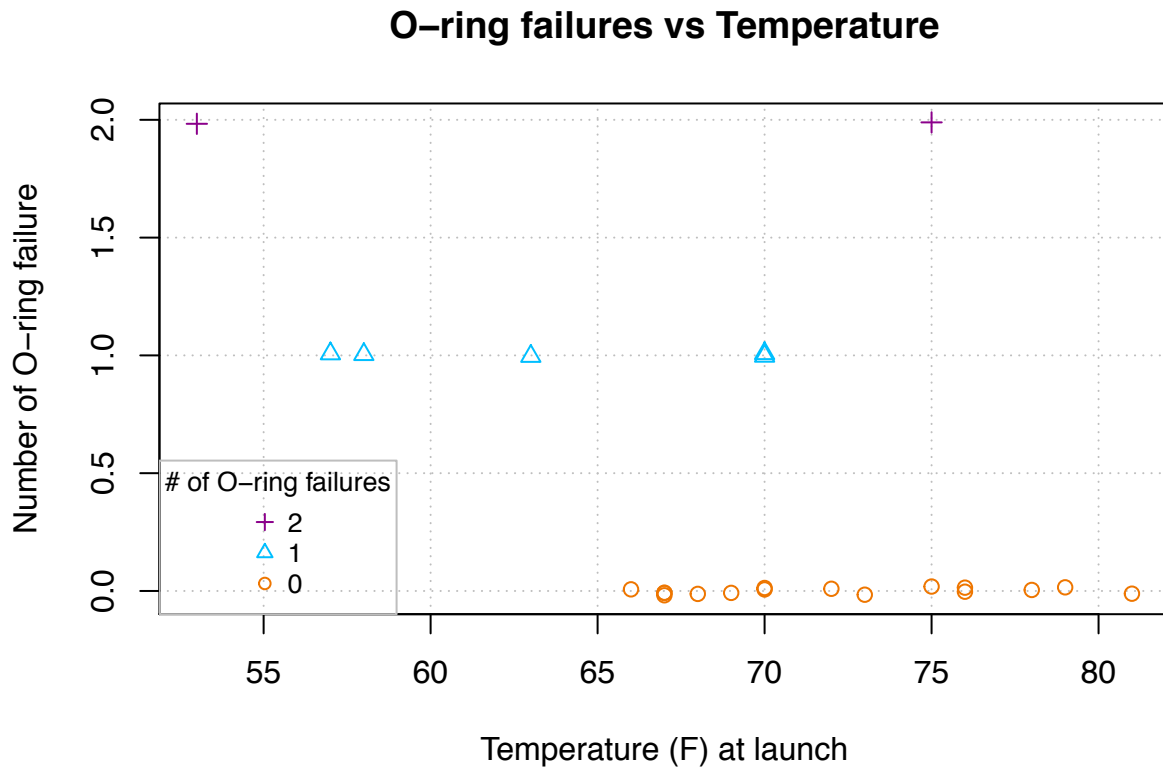

2.3.2 Bivariate Relationship : O.ring failure and Temperature

- The following scatterplot and boxplot illustrates relationship between number of Primary O-ring failures with temperature. Despite the data limitation that we have relatively few observations with O-ring failures, it's obvious that these failures tend to happen when temperature is lower than the overall mean temperature at around 70F. If we regress O-ring failure on temperature, we should expect a negative coefficient.

```
#plot(df$O.ring~df$Temp, pch = 3, xlim = c(30,90), )
plot(df$Temp, jitter(df$O.ring,0.1)
     ,main = "O-ring failures vs Temperature"
     ,pch = ifelse(df$O.ring == 0
                   ,1
                   ,ifelse(df$O.ring == 1
                           ,2 ,3))
     ,ylab = "Number of O-ring failure", xlab = "Temperature (F) at launch"
     ,col = ifelse(df$O.ring == 0
                   ,"darkorange2"
                   ,ifelse(df$O.ring == 1
                           ,"deepskyblue", "darkmagenta")),
     panel.first = grid(col = "gray", lty = "dotted"))

legend("bottomleft", legend = c("2","1","0"),
```

```
        col = c("darkmagenta", "deepskyblue", "darkorange2"),
        pch = c(3,2,1),
        box.col = "gray",
        cex=.8,
        title = "# of O-ring failures")
```

O–ring failures vs Temperature
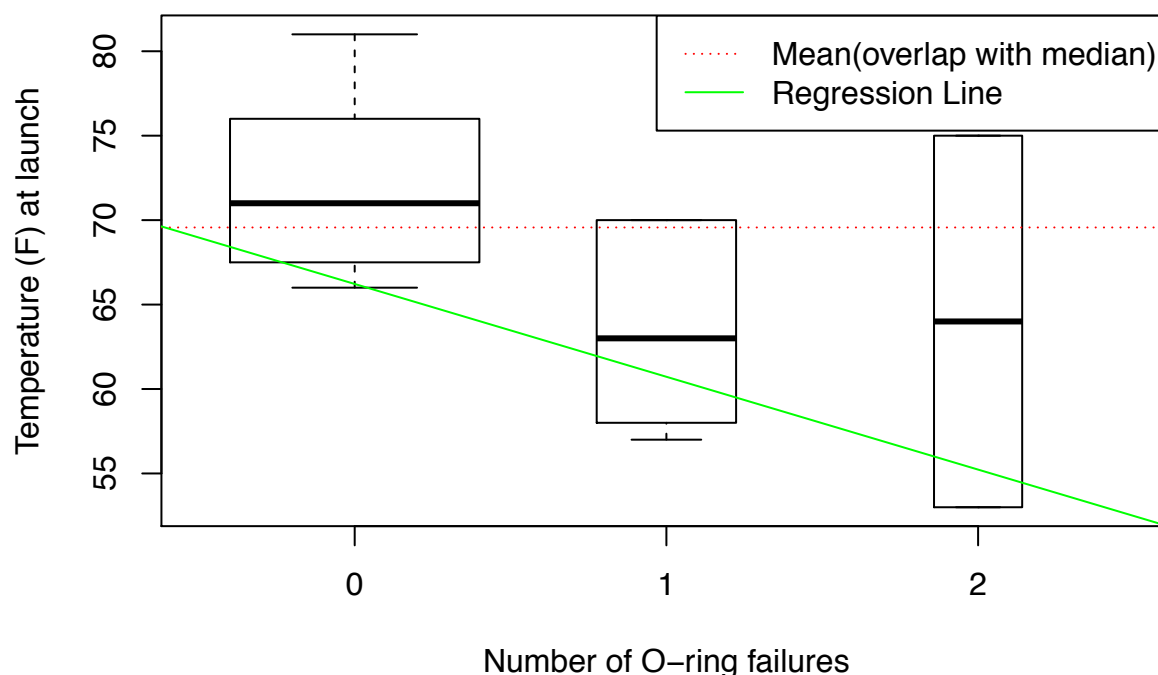


```
boxplot(df$Temp ~ df$O.ring
        , main = "O-ring failures vs Temperature"
        , xlab = "Number of O-ring failures"
        , ylab = "Temperature (F) at launch", varwidth = T)

abline(h = mean(df$Temp), col = "red", lty = "dotted")
abline(lm(df$Temp ~ df$O.ring),col = "green")

legend("topright", c("Mean(overlap with median)", "Regression Line"), col = c("red", "green"), lty = c(
```

O–ring failures vs Temperature

2.3.3 Bivariate Relationship : O.ring failure and Pressure

- The following contingency table illustrates relationship between number of Primary O-ring failures with pressure. We see that of the 23 launches, 7 experienced O-ring failures. Out of these 7 launches, 6 took place when pressure was held at 200psi and 1 was held when pressure was held at 50psi. At first glance this may suggest a positive effect of pressure on O-ring failures, but notice that out of the 15 launches held at 200psi, 9 launches experienced no failures. Therefore, we can't develop a strong intuition of pressure's effect on O-ring failures yet.

```
xtabs(~ df$O.ring + df$Pressure)
```

```
##          df$Pressure
## df$O.ring 50 100 200
##         0  5   2   9
##         1  1   0   4
##         2  0   0   2
```

- The following box plot concurs with the observations and intuitions from the contingency table.

```
boxplot(df$O.ring ~ df$Pressure
        , main = "Pressure vs O-ring failures"
        , xlab = "Number of O-ring failures"
        , ylab = "Pressure (psi) at launch", varwidth = T)

abline(h = mean(df$O.ring), col = "red", lty = "dotted")
abline(h = median(df$O.ring), col = "blue", lty = "dotted")
abline(lm(df$O.ring ~ df$Pressure),col = "green")
```

```
legend("topleft", c("Mean", "Median", "Regression Line"), col = c("red", "blue","green"), lty = c("dotte
```

Pressure vs O–ring failures
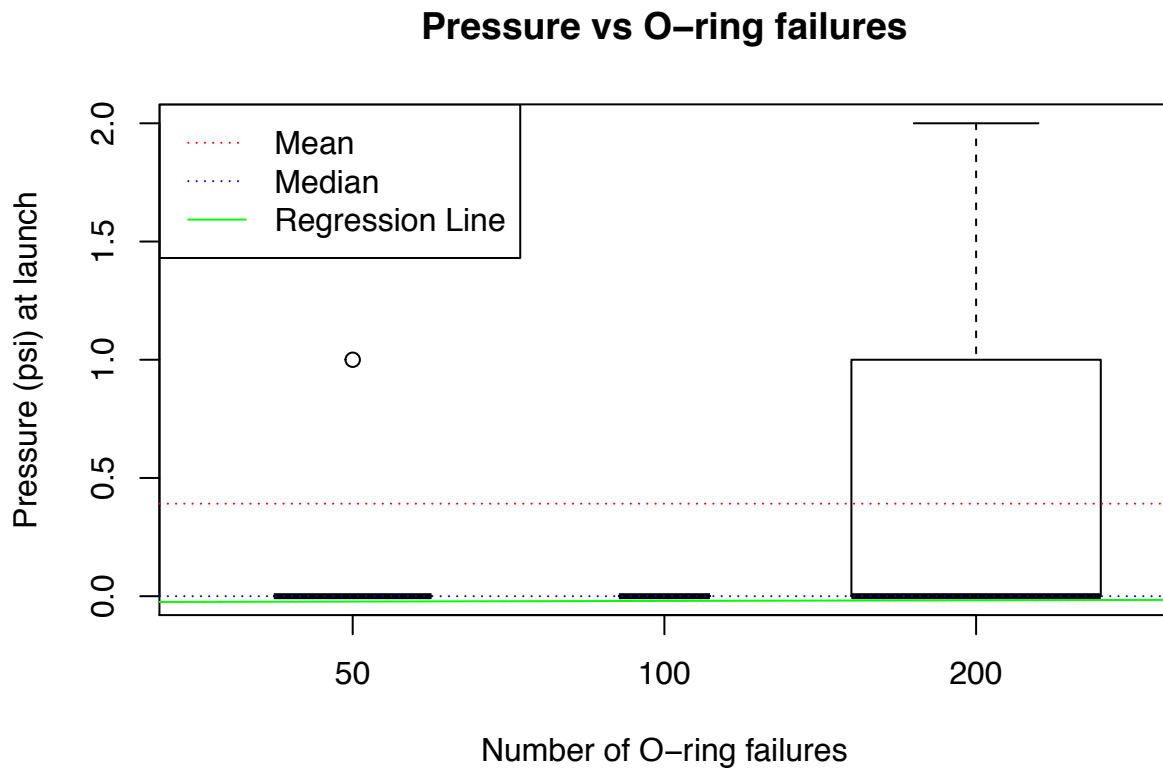
2.3.4 Bivariate Relationship : Temperature and Pressure

- The following plot illustrates joint distribution of launches on temperature against pressure. Launches which experinced zero, one or more O-ring failures are annotated in different colors. Notice that at pressure 200psi and temperature below 65F, all four launches experienced O-ring failures. We suspect that an interaction effect is at play and will test its significance in section 4.

```
plot(df$Pressure, jitter(df$Temp)
     ,main = "O-ring failures: Temperature vs Combustion Pressure"
     ,pch = ifelse(df$O.ring == 0
                   ,1
                   ,ifelse(df$O.ring == 1
                           ,2 ,3))
     ,ylab = "Temperature (F) at launch", xlab = "Combustion pressure (psi)"
     ,col = ifelse(df$O.ring == 0
                   ,"darkorange2"
                   ,ifelse(df$O.ring == 1
                           ,"deepskyblue", "darkmagenta")),
     panel.first = grid(col = "gray", lty = "dotted"))

#boxplot(df$Temp ~ df$Pressure, varwidth = T)

abline(h = mean(df$Temp), col = "red", lty = "dotted")
```
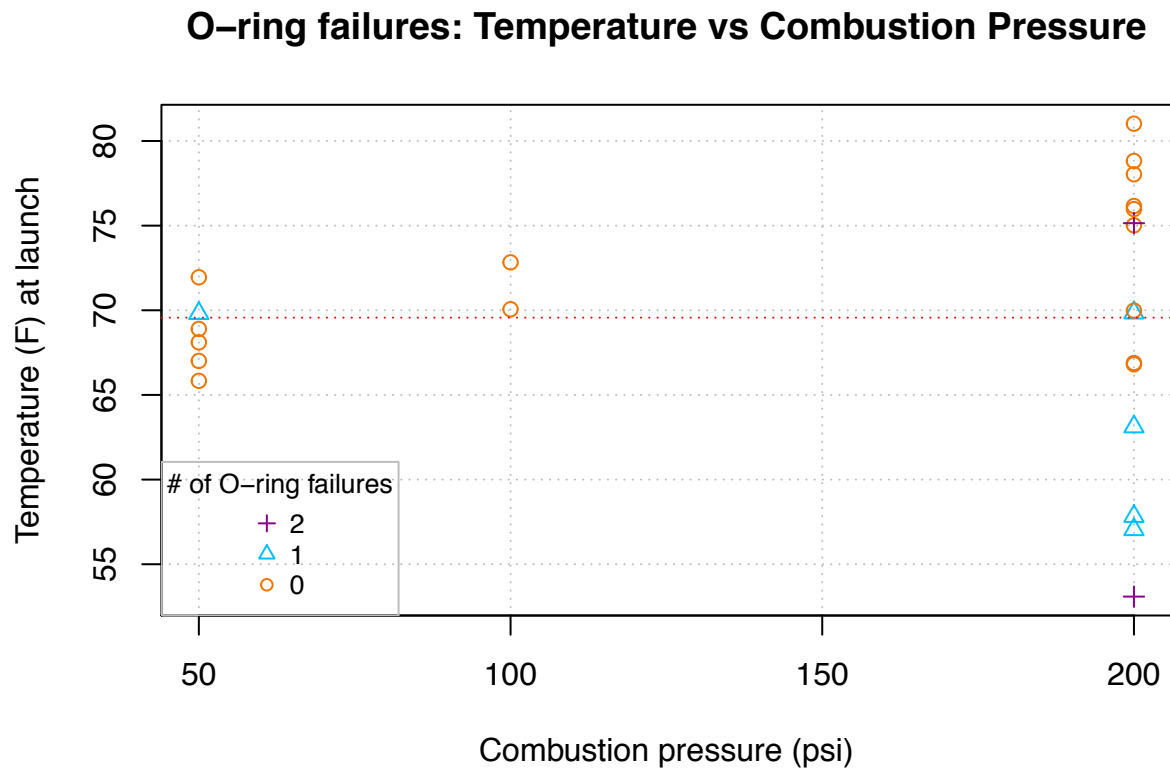
```
legend("bottomleft", legend = c("2","1","0"),
       col = c("darkmagenta", "deepskyblue", "darkorange2"),
       pch = c(3,2,1),
       box.col = "gray",
       cex=.8,
       title = "# of O-ring failures")
```

O–ring failures: Temperature vs Combustion Pressure

2.4 Note on Random Sampling

- Using the EDA above, we also developed some intuition about sampling of the data. Since some variables (such as pressure) are obviously manipulated and we are interested in causal inference, we try to understand the observations as an experiment dataset. We may assume that our participants are the 23 launches, our independent variables of interest are temperature and pressure. We are trying to associate exposure to temperature and pressue with an increase or decrease in occurence of primary O-ring failure.

  - No random sampling: We may consider the population as all launches with solid rocket motors installed with O-rings. The challenger launches are not a random selection because it's associated with the same rocket. We may limit the scope of study to be specific about solid rocket motors with the challenger's design. In this case maturation, that some of the challenger's parts have aged over the history of launches, can threaten internal validity of our experiment.

  - Not random assignment: launches were assigned to optimum temperature and pressure that avoid failure, not randomly to various conditions.

  - Semi-active Independent variables:"Pressure" was manipulated for optimum launching condition,

14

not to study its effects. The disproportionate number of observations at 200psi, 100psi and 50psi do not serve our study well. "Temperature" has some natural variation but again was not actively manipulated for our study, fewer observations at lower temperatures do not serve our study well again.

2.5 Modeling Choices

2.5.1. Inappropriateness of OLS regression

- In the context of regression modeling, we have examined the following assumptions in the above EDA:
  - Linearity in parameters: We have already identified a moderate linear relationship between O-ring failure and temperature. This is not a strong assumption because we still have flexibility in explanatory variable specification.
  - Random Sampling: The observations at best qualify as associative non-experiment. We should keep in mind that conclusion from this study cannot be generalized to other rockets.
  - Family of distribution: Our response variable of interest is the expected number of O-ring failure, which is proportional to the expected probability of failure. It is bounded between 0 and 1. It also resembles a binomial distribution of n = 6 and probability of success of 0.065 (mean of observations).
  - We acknowledge that OLS regression is not appropriate because its estimates are unbounded. That is, the estimated probability can be lower than 0 or great than one, equivalent to estimating less than 0 or more than 6 O-ring failures. Still, we construct an OLS model for demonstration purposes and evaluate the model's other shortcomings.

```
O.ring.lm = lm(O.ring ~ Temp + Pressure, data = df)

suppressWarnings(stargazer::stargazer(O.ring.lm
                    ,type = "text", title = "OLS: Temp and Pressure"
                    ,star.cutoffs = c(0.05, 0.01, 0.001)))
```

```
##
## OLS: Temp and Pressure
## =================================================
##                          Dependent variable:
##                      ----------------------------
##                                O.ring
## -------------------------------------------------
## Temp                          -0.049**
##                                (0.017)
##
## Pressure                       0.003
##                                (0.002)
##
## Constant                       3.330*
##                                (1.188)
##
## -------------------------------------------------
## Observations                      23
## R2                              0.354
## Adjusted R2                     0.290
## Residual Std. Error      0.553 (df = 20)
## F Statistic           5.490* (df = 2; 20)
## =================================================
```

```
## Note:                     *p<0.05; **p<0.01; ***p<0.001
```

- The positive sign of "Temp"'s coefficient and negative sign of "Pressure"'s coefficient agree with our intuition developed from the EDA. "Temp" is the only statistically significant coefficient with value of $-0/049$. An increase of 7.057F, a standard deviation in temperature, is estimated to decrease the expected number of O-ring failure by 0.343. The coefficient has some practical significance.

- Holding pressure at 200psi, once temperature drop to around -45F, the estimated number of O-ring failure raise to above 6, that is more O-rings than the rocket obtain. Similar result happends wheb we hold temperature at 31F and raise pressure to around 1500psi.
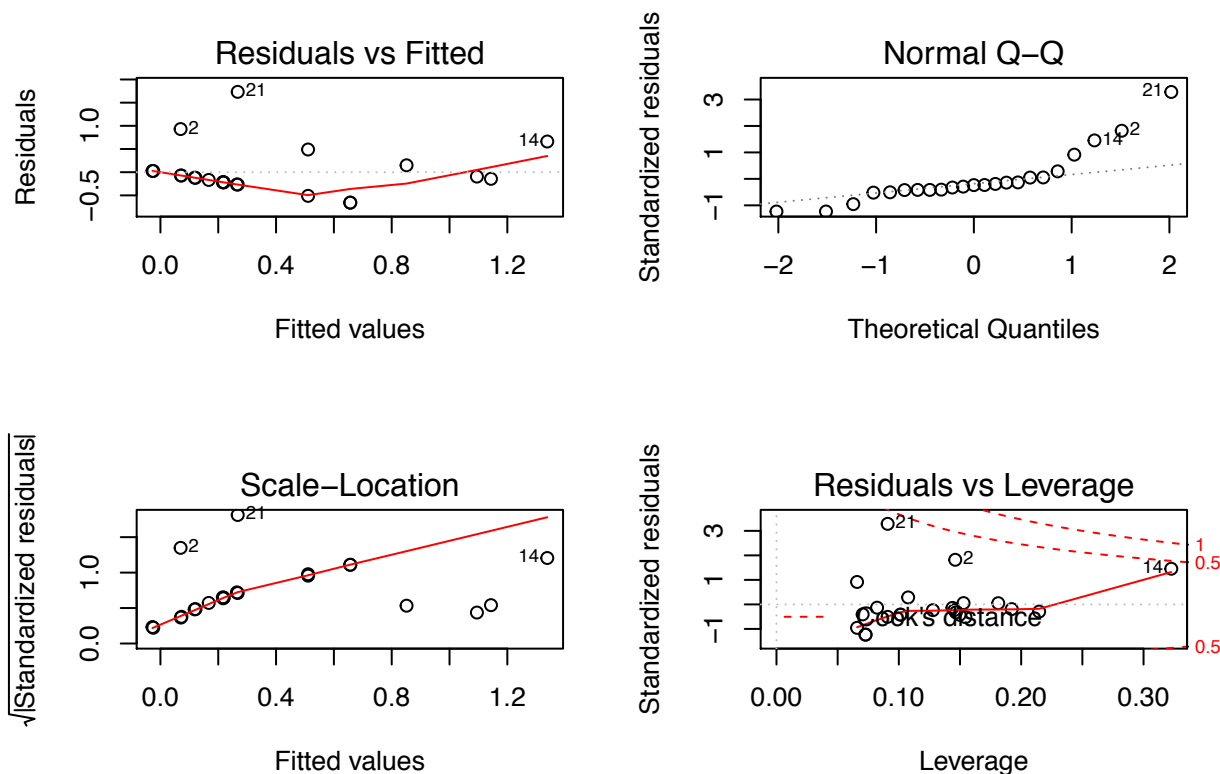
```
predict(object = O.ring.lm, newdata = data.frame(Temp = -45, Pressure = 200))
```

```
##        1
## 6.107895
```

```
predict(object = O.ring.lm, newdata = data.frame(Temp = 31, Pressure = 1500))
```

```
##        1
## 6.230014
```

```
par(mfrow = c(2, 2))
plot(O.ring.lm)
```


- From the residual plots above, we know that several of our OLS assumptions are broken.

  - Zero-conditional mean violation: the dip of the Loess curve on our residual vs fitted values plot shows a systematic violation of zero-conditional mean. This means our explanatory variables can be mis-specified and our coefficient estimates are biased. We can consider our model associative instead of causal and resort to exogeneity. However, given our small sample size, our estimator is

not consistent either.

- Heteroskedasticity: the scale location plot shows an uneven distribution of residuals with an upward take-off. This means variance of our residuals is increasing with our variables. Our OLS estimator is not very efficient.

- Normality of Residuals: the Normal Q-Q plot shows a handful of data points taking off from the normal distribution curve. This is expected because our reponse variable distribution is also positively skewed. Our coefficient estimates also mirror this distribution. This means we can't effectively conduct hypothesis testing on our coefficient which predicate on normality of coefficients.

- Outliers: we have 3 observations (2, 14, 21) close to the cook's distance of 0.5 curve. Observation 2 and 14 has a standardized residual over 1, while observation 21 has a standardized residual well over 3. A closer examination of these observations reveals that observations 2 and 21 experienced O-ring one or more failure above the mean temperature at 69.57F. This is contrary to our intuition about temperature's effect on odds of O-ring failure. This also means that our OLS model is representing our data poorly with problem of over-dispersion. We take note of these two outliers to see if they are still outstanding in improved models.

```
df[c(2,14,21),]
```

```
##    Flight Temp Pressure O.ring Number O.failed.binary
## 2       2   70       50      1      6               1
## 14     14   53      200      2      6               1
## 21     21   75      200      2      6               1
```

Question 4

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.

Answer 4a)

This assumption is needed in the binomial logistic regression model because:

- Maximum Likelihood is used to estimate parameters of our model. In its working mechanism, a joint probability distribution is derived by multiplying together binomial probability mass function of each of our observations as if they were independent. Numerical procedures search for coefficients which maximizes:

$$L(\beta_0, ..., \beta_p | y_1, ..y_n) = \prod_{i=1}^{n} \pi_i^{y_i} (1 - \pi_i)^{(1-y_i)}$$

- It is necessary that each observation is independent for us to multiply(join) all pmfs, i.e. $\pi_i^{y_i}(1-\pi_i)^{(1-y_i)}$ together.

- However, this assumption is not always a given. In the Challenger's case, it is not realistic because the O-rings are manufactured by the same company under the same industrial process, same design and potentially from same patch of raw material. O-rings which belong to the same launch is also subjected to same launching and testing conditions other than temperature and pressure.

- If this assumption is violated, our MLE estimates will be biased. The author's subsequent, binary model used a reponse variable of simply "any failure"/"no failure" which eliminated this concern. In a side-by-side visual comparison, the two models were close with the binomial model fitting the observations better.

- Note that when the author applies the same modeling procedure to the nozzle, instead of field O-ring data, the binomial fit was poorer than the binary model. Indiciating the draw back of binomial model for dependent data.

4b) Estimate the logistic regression model using the explanatory variables in a linear form

Answer 4b)

- Below we are providing two logistic regression models for binomial and binary response. For binomial, we feed in proportion of O-ring failure at each launch as a response variable. For binary, we feed in 0 or 1 to indicate whether O-ring failure happened during each flight.

$$log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_t Temperature + \beta_s Pressure$$

4b) 1.Binomial Logistic Regression Model

```
df$O.failed.binomial = df$O.ring / df$Number

O.failed.binomial.glm = suppressWarnings(glm(formula = O.failed.binomial ~ Temp + Pressure,
                         family = binomial(link = logit), data = df))

summary(O.failed.binomial.glm)
```

```
##
## Call:
## glm(formula = O.failed.binomial ~ Temp + Pressure, family = binomial(link = logit),
##     data = df)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.42299  -0.26267  -0.21671  -0.06634   0.95601
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.520195   8.540918   0.295    0.768
## Temp         -0.098297   0.109959  -0.894    0.371
## Pressure      0.008484   0.018806   0.451    0.652
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4.0384  on 22  degrees of freedom
## Residual deviance: 2.7576  on 20  degrees of freedom
## AIC: 9.2286
##
## Number of Fisher Scoring iterations: 6
```
```
logLik(O.failed.binomial.glm)
```

```
## 'log Lik.' -1.614316 (df=3)
```

- Point estimations of the coefficients in our binomial model agree with that of the author's. We acknowledge that our standard errors are different from the author's and a quasi binomial regression

model would have given closer estimates. We provide a residual analysis with plots shortly to address this.

- – Temperature (-0.098) has a negative sign, suggesting an increase of temperature decreases expected odds of O-ring failure. The estimated odds of O-ring failure changes by 0.5 times ,equivalent to decrease by around 50%, with one standard deviation of increase (7.057F) in temperature , holding pressure constant.

$$OR = \frac{Odds_{T+c}}{Odds_T} = \frac{e^{\beta_0 + \beta_T(T+c) + \beta_S S}}{e^{\beta_0 + \beta_T T + \beta_S S}} = e^{c\beta_T} = e^{7.057*(-0.098)} = 0.50$$

- Pressure (0.008) has a positive sign, suggesting an increase of pressure increases expected odds of O-

$$OR = \frac{Odds_{S+c}}{Odds_S} = \frac{e^{\beta_0 + \beta_T T + \beta_S(S+C)}}{e^{\beta_0 + \beta_T T + \beta_S S}} = e^{c\beta_S} = e^{68.221*(0.008)} = 1.784$$

- Regarding the significance of our estimates, Wald Test ueses the hypotheses:

  - – Ho: the estimated coefficient is zero.
  - – Ha: the estimated coefficient is not zero.

$$Z_o = \frac{\hat{\beta}_r - 0}{\sqrt{Var(\hat{beta}_r)}}$$

- If we "invert" the test statistic, we get Wald Confidence Interval:

$$\hat{\pi} \pm Z_{1-\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

where coefficients are linked to probability estimates by:

$$\hat{\pi} = \frac{e^{\beta_0 + \beta_t Temperature + \beta_s Pressure}}{1 + e^{\beta_0 + \beta_t Temperature + \beta_s Pressure}}$$

```
confint.default(object = O.failed.binomial.glm, level = 0.95)
```

```
##                   2.5 %      97.5 %
## (Intercept) -14.21969641 19.26008566
## Temp         -0.31381188  0.11721838
## Pressure     -0.02837439  0.04534244
```

- From the default regression output, with a 95% confidence, none of the coefficients are significant using Wald statistics. Coefficient estimate for temperature has a p value of 0.371 and for pressure is 0.652, both of which are greated than 0.05.This raises questions regarding the usefulness of these coefficients. Wald Confidence Intervals for our estimated coefficients both covers zero. Agreeing that the null hypothesis cannot be rejected.

- Note that with our small sample size of n = 23, Wald Test is not appropriate because it assumes an asymptotic distribution to work and it's also too liberal. Also, the discreteness of binomial distribution generate fluctuating true confidence levels, especially when estimated reponses are close to 0 or 1.

- Below we conduct a residual analysis for the model:
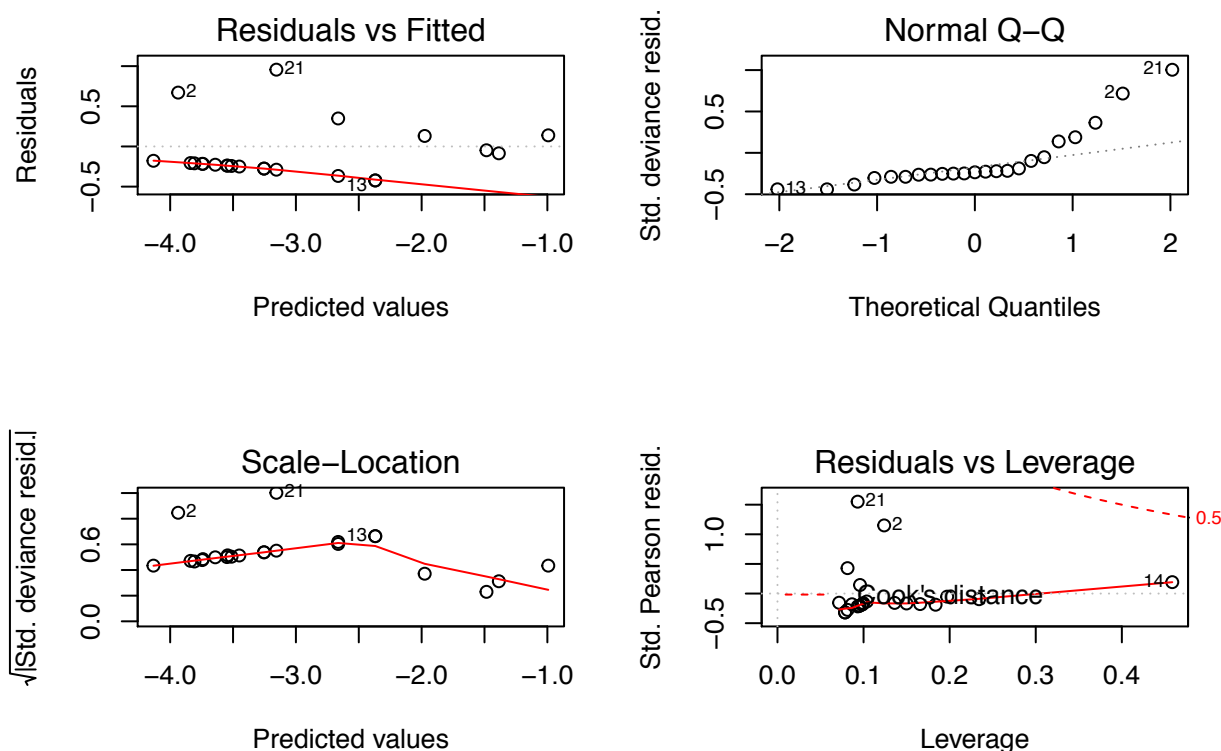
```
par(mfrow = c(2, 2))
plot(O.failed.binomial.glm)
```

- From the residual plots above, we notice the following features of our residuals:

  - Zero-conditional mean violation: the mean of our residuals trends downwards on the residual vs fitted values plot. 4 observations picked up at the tail but weren't enough to pick up the trend. This systematic departure suggests that our variables can be mis-specified and their estimation can be biased. We can consider our model associative instead of causal and resort to exogeneity. However, given our small sample size, our estimator is not consistent either.

  - Heteroskedasticity: compared to the OLS model, this scale location plot shows more stable standardized residuals with a raise in the middle so it's harder to judge visually. We use the Breusch Pagan Test below to test for heteroskedasticity.

  - H_o : variance of our residual changes with fitted values

  - H_a : variance of our residual does not change with fitted values

  - With a p-value of 0.749 we failed to reject the null hypothesis. However, our small sample size can also contribute to this insignificance. The Breusch Pagan test regresses our residuals on fitted values to estimate significance of the resultant coefficient. With our small sample size there is large variability in this estimated coefficient, therefore harder for it to establish statistical significance. We recommend using Heteroskedasticity Robust Standard Errors provided below.

```
lmtest::bptest(O.failed.binomial.glm)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  O.failed.binomial.glm
## BP = 0.57721, df = 2, p-value = 0.7493
```

```r
sqrt(diag(vcovHC(O.failed.binomial.glm)))
```

```
## (Intercept)        Temp    Pressure
## 3.937970565 0.049519653 0.008969315
```

- Normality of Residuals: Same as the OLS model the Normal Q-Q plot shows a handful of data points takin

  - Ho: our standardized residuals have the same distribution as standard normal distribution
  - Ha: our standardized residuals does not have the same distribution as standard normal distribution

  - This means that Wald test is not appropriate to test significance of our ceofficients because they p

```r
shapiro.test(O.failed.binomial.glm$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  O.failed.binomial.glm$residuals
## W = 0.49346, p-value = 7.427e-08
```

- Outliers: Compared to the OLS residual plots, only observations 2 and 21 stand out on our residuals p

```r
O.failed.binomial.glm.quasi = glm(formula = O.failed.binomial ~ Temp + Pressure,
                        family = quasibinomial(link = logit), data = df)
par(mfrow = c(2, 2))
plot(O.failed.binomial.glm.quasi)
```
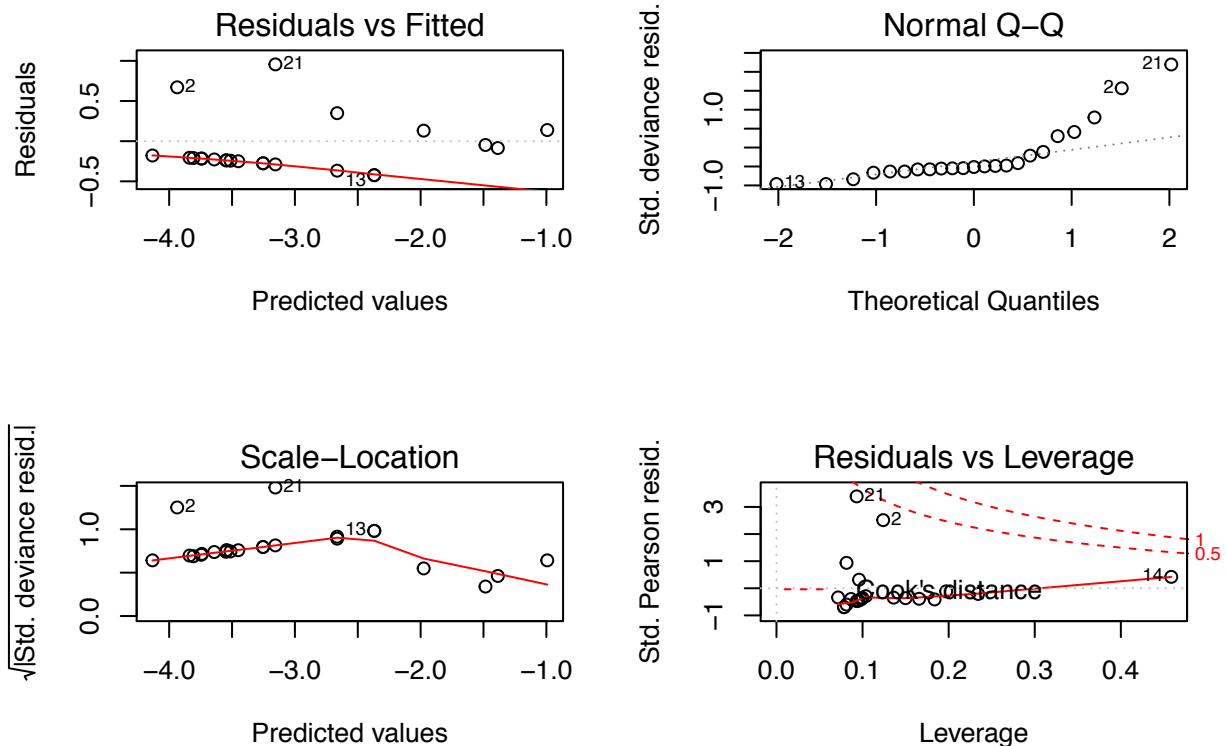

- We take note of the regression outputs to assess goodness of fit:

- Residual deviance is 2.7576. A model with smaller residual deviance is considered to be a better fit. However, like R^2, residual deviance is always increasing with addition of new variables, even with random noises.

- Log Likelihood is -1.6143. A model with higher log likelihood is considered to be a better fit. Later, Likelihood Ratio Tests can compare two models which differ by nesting variables as polynomial or interaction terms in its specification.

- AIC is 9.2286. A model with smaller AIC is considered to be a better fit, assuming the dataset and family of model is the same.

4b) 2.Binary Logistic Regression Model

```
O.failed.binary.glm = glm(formula = O.failed.binary ~ Temp + Pressure ,
                          family = binomial(link = logit), data = df)

summary(O.failed.binary.glm)
```

```
##
## Call:
## glm(formula = O.failed.binary ~ 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
```

```
logLik(O.failed.binary.glm)
```

```
## 'log Lik.' -9.391047 (df=3)
```

- With the binary model, signs of the coefficients agrees with our binomial model.

- Temperature (-0.229) has a negative sign, suggesting a increase of temperature decreases expected odds of O-ring failure. The estimated odds of O-ring failure changes by 0.199 times ,equivalent to decrease by around 80%, with one standard deviation of increase (7.057F) in temperature , holding pressure constant. The estimated effect is more dramatic than the binomial model.

$$OR = \frac{Odds_{T+c}}{Odds_T} = \frac{e^{\beta_0 + \beta_T(T+c) + \beta_S S}}{e^{\beta_0 + \beta_T T + \beta_S S}} = e^{c\beta_T} = e^{7.057*(-0.229)} = 0.199$$

- Pressure (0.010) has a positive sign, suggesting an increase of pressure increases expected odds of O-ring failure. The estimated odds of O-ring failure changes by 2.033 times ,equivalent to increase by around 103%, with one standard deviation of increase (68.221psi) in temperature, holding pressure constant. The estimated effect is also more dramatic than the binomial model.

$$OR = \frac{Odds_{S+c}}{Odds_S} = \frac{e^{\beta_0 + \beta_T T + \beta_S (S+C)}}{e^{\beta_0 + \beta_T T + \beta_S S}} = e^{c\beta_S} = e^{68.221*0.010} = 2.033$$

- Regarding the significance of our estimates, Wald Test mechanism is the same as above:
    - Ho: the estimated coefficient is zero.
    - Ha: the estimated coefficient is not zero.

```
confint.default(object = O.failed.binary.glm, level = 0.95)
```

```
##                    2.5 %       97.5 %
## (Intercept) -1.728742490 28.31346153
## Temp        -0.444244003 -0.01309790
## Pressure    -0.007198764  0.02799905
```

- With a 95% confidence, estimate for the coefficient of temperature is statistically significant with p-value 0.038, but that of pressure is statistically insignificant with p-value 0.247. The confidence interval for temperature just clear from zero and that of pressure covers zero. Caveats of Wald Test (discussed above) still applies.

- Below we conduct a residual analysis for the model:
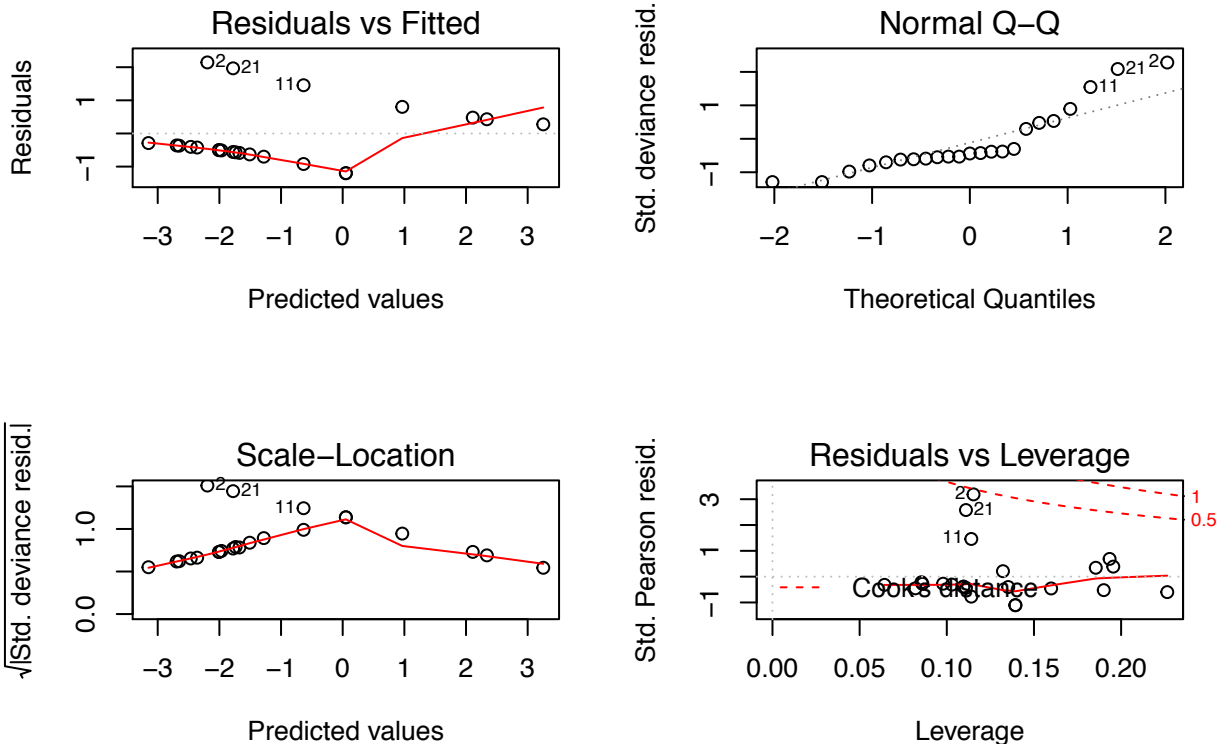
```
par(mfrow = c(2, 2))
plot(O.failed.binary.glm)
```

- From the residual plots above, we notice the following features of our residuals:

  - Zero-conditional mean violation: the mean of our residuals trends downwards on the residual vs fitted values plot. 4 observations picked up at the tail but weren't enough to pick up the trend. This systematic departure suggests that our variables can be mis-specified and their estimation can be biased. We can consider our model associative instead of causal and resort to exogeneity. However, given our small sample size, our estimator is not consistent either.

  - Heteroskedasticity: similar to the binomial model, this scale location plot shows a raise in residuals values in the middle so it's hard to judge constant variance visually. We use the Breusch Pagan Test again to test for heteroskedasticity.

  - H_o : variance of our residual changes with fitted values

  - H_a : variance of our residual does not change with fitted values

  - With a p-value of 0.958 we failed to reject the null hypothesis. However, our small sample size can also contribute to this insignificance. We recommend using Heteroskedasticity Robust Standard Errors provided below.

```
lmtest::bptest(O.failed.binary.glm)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  O.failed.binary.glm
## BP = 0.086582, df = 2, p-value = 0.9576
```

```
# White standard errors
# lmtest::coeftest(O.failed.binary.glm, vcov = vcovHC)
sqrt(diag(vcovHC(O.failed.binary.glm)))
```

```
## (Intercept)        Temp     Pressure
##  5.98639835  0.08611590  0.01101542
```

- Normality of Residuals: Similar to the binomial model the Normal Q-Q plot shows a handful of data poi

```
stu_residuals = MASS::studres(O.failed.binary.glm)
hist(stu_residuals, col = rgb(1, 0, 0, .5), ylim = c(0,1),
main = "Studentized Residuals", xaxt = "n", breaks = "FD", freq = F)
norm.dis.x = rnorm(n = 100, mean = 0, sd = 1)
hist(norm.dis.x, add = T, freq = F , col = rgb(0,0,1,.5))
axis(1, at = seq(-6,5,1))

legend("topright" , cex = 1, title = "legend", legend = c("studentized residual", "standardized normal
, fill = c(rgb(1, 0, 0, .5),rgb(0,0,1,.5)))
```

Studentized Residuals
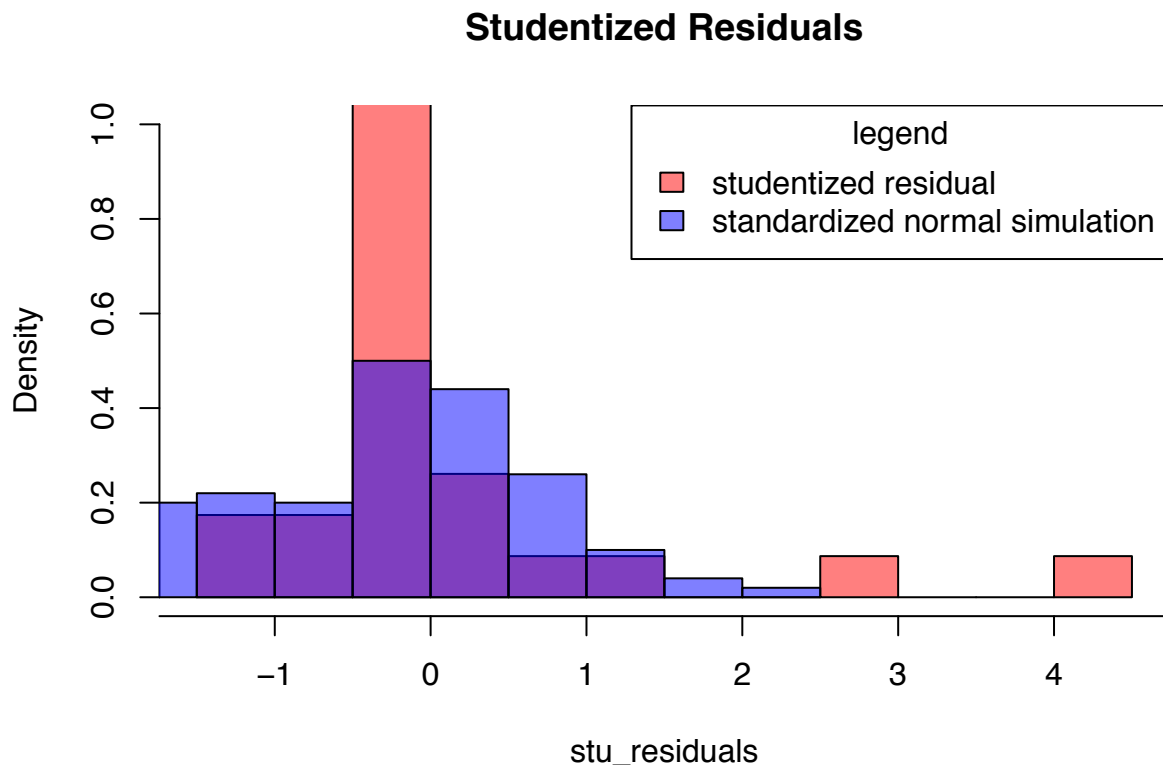
- Shapiro test confirms non normality even with a sample size of 23.

    - Ho: our standardized residuals have the same distribution as standard normal distribution
    - Ha: our standardized residuals does not have the same distribution as standard normal distribution

    - This means that Wald test is not appropriate to test significance of our ceofficients because the

```r
shapiro.test(O.failed.binary.glm$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  O.failed.binary.glm$residuals
## W = 0.64539, p-value = 2.959e-06
```

- Outliers: With the binary logistic model, observation 2 and 21 behave more extremely than the Binomia

- We take note of the residual deviance (18.782), log likelihood (-9.391) and AIC (24.782). Unfortunately, we can't use the AIC to compare these two models because the response variable is different (binary vs binomial). LRT Anova is not appropriate either because the specification difference between the models is not nesting.

- To compare the two models side by side, we have constructed the plots of expected number of accidents vs temperature(key explanatory variable of interest) for the two models, holding pressure at it mean value (152.17). Generally, both models related probability of O-ring failure to temperature and pressure by:

$$\pi_i = \frac{e^{\beta_0 + \beta_T T + \beta_S S}}{1 + e^{\beta_0 + \beta_T T + \beta_S S}}$$

```r
psi = mean(df$Pressure)

# base plot of observations
plot(x = df$Temp, y=df$O.ring, xlab = "Temperature in F",
     ylab = "Expected Number of Incidents",
     panel.first = grid(col = "gray", lty = "dotted"), ylim = c(0,6), xlim = c(0,90),
     main = "Field-joint O-ring Distress Data: Binomial vs Binary Logit")

curve(expr = predict(object = O.failed.binomial.glm,
      newdata = data.frame(Temp= x, Pressure = psi), type = "response")*6,
      col = "red", xlim = c(0,90), add = T)

curve(expr = predict(object = O.failed.binary.glm,
      newdata = data.frame(Temp= x, Pressure = psi), type = "response")*6,
      col = "blue", xlim = c(0,90), add = T)

# Legend
legend(x = 0, y = 2, legend = c("Binomial Logit model", "Binary Logit model"),
       lty = c("solid", "solid"), col = c("red", "blue"), bty = "n")
```

Field–joint O–ring Distress Data: Binomial vs Binary Logit
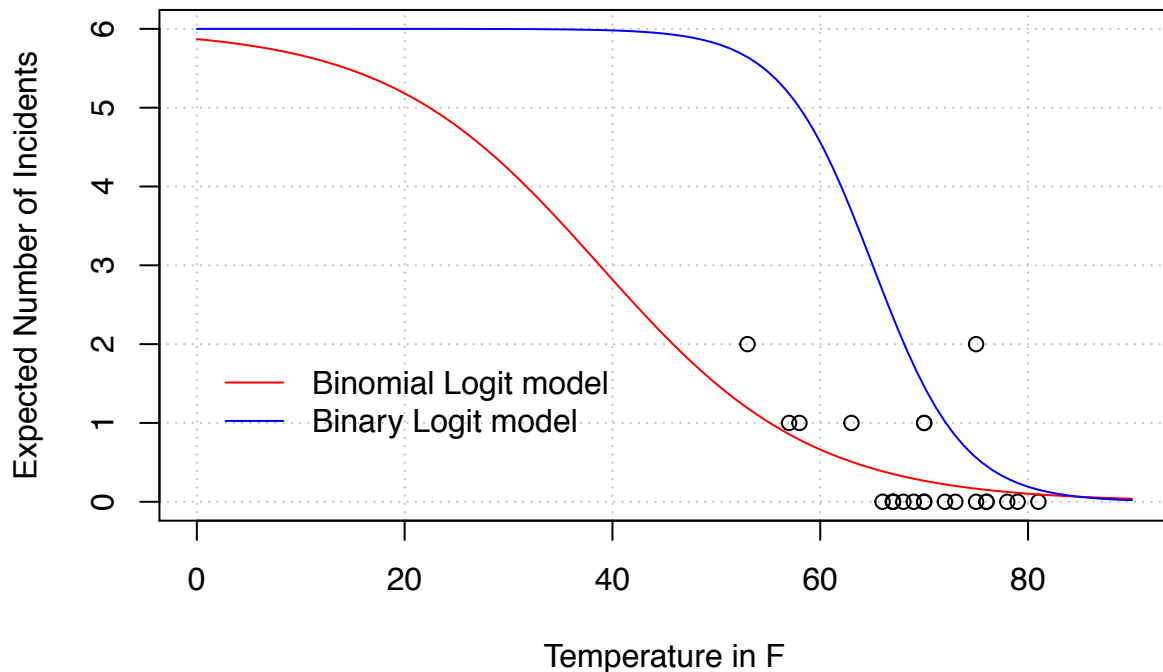
- At 152.17 psi, both models agree that a decrease in temperature increases the expected number of incidents. However, the difference between the two fits are quite different. The binary model and a steeper fitted curve and estimates all O-rings to fail when temperature is below 40F. While the binomial

model seem to fit the observations better. It has a genlter fitted curve and estimates all O-rings to fail at some temperature below 0 F.

- We are not imposing any Wald Confidence Intervals here for its inappropriateness discussed above.

4c) Perform LRTs to judge the importance of the explanatory variables in the model.

Answer 4c)

- Below we will use Likelihood Ratio Test (LRT) to assess importance of explanatory variables in the above two models. LRT assess whether the inclusion of a particular explanatory variable changes the overall predicting power of the model significantly, by comparing Maximum Likelihood values between the restricted and unrestricted model. It tests statistic $-2log(\Lambda)$, is used as a measure of deviance between models.

- Compared to Wald Test, LRT's test statistic is more stable when estimates are close to zero or one. Although both tests suffer from a small sample size, LRT is more appropriate.

- An LRT is performed in each model for each coefficient below with theses hypotheses:

  - Ho: $\beta_r = 0$

  - Ha: $\beta_r \neq 0$

  - In each test, our restricted model hold all other variable present but the variable of interest. We use the Anova() function thus the test does not run in a sequential manner.

```
LR.binomial = suppressWarnings(car::Anova(O.failed.binomial.glm))
LR.binomial
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: O.failed.binomial
##          LR Chisq Df Pr(>Chisq)
## Temp      0.86397  1     0.3526
## Pressure  0.25678  1     0.6123
```

```
PLR.binomial = suppressWarnings(confint(object = O.failed.binomial.glm, level = 0.95))
PLR.binomial
```

```
##                   2.5 %     97.5 %
## (Intercept) -17.84080488 23.2811240
## Temp         -0.37938750  0.1159626
## Pressure     -0.02191396         NA
```

- For our binomial logit model, comparing Wald Test results head to head with LRT results:

  - Temp ~~remain~~ remain insignificant, with Wald p-value 0.371 being larger than LRT p-value 0.3526309

  - Wald confidence interval for Temp (-0.31381188, 0.11721838) is slightly narrower (by difference of 0.064F) than Profile Likelihood Interval (-0.37938750, 0.1159626)

  - Pressure remain insignificant, with Wald p-value 0.652 being larger than LRT p-value 0.6123437

  - Wald confidence interval for Pressure is (-0.02837439, 0.04534244). Profile Likelihood Interval cannot be computed (NA result).

  - Overall, our LRT results agrees with Wald Test results.

```
LR.binary = car::Anova(O.failed.binary.glm)
LR.binary
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: O.failed.binary
##          LR Chisq Df Pr(>Chisq)
## Temp       7.7542  1   0.005359 **
## Pressure   1.5331  1   0.215648
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
PLR.binary = confint(object = O.failed.binary.glm, level = 0.95)
PLR.binary
```

```
##                     2.5 %      97.5 %
## (Intercept)   0.973692095 33.09799473
## Temp         -0.518222024 -0.05703940
## Pressure     -0.005751428  0.03226285
```

- For our binary logit model, comparing Wald Test results head to head with LRT results:

    - Temp becomes considerably significant, with Wald p-value 0.038 being larger than LRT p-value 0.0053587. Wald confidence interval for Temp (-0.444244003, -0.01309790) is slightly narrower (by difference of 0.030 F) than Profile Likelihood Ratio Interval (-0.518222024, -0.05703940).

    - Pressure remain insignificant, with Wald p-value 0.247 being larger than LRT p-value 0.2156479. Wald confidence interval for Pressure (-0.007198764, 0.02799905) is again narrower (by difference of 0.003 psi) than Profile Likelihood Ratio Interval (-0.005751428, 0.03226285). Both intervals cover zero, agreeing that pressure is statistically insignificant.

- Overall, results from Wald Test and LRT on our two models agree. With Wald Test p values more conservative and Profile Likelihood Intervals slightly wider. However, we should note either test suffer from a small sample size. Thus, the author's approach of bootstrapping for confidence intervals seem more appropriate.

4d) The authors chose to remove Pressure from the model based on the LRTs. Based on your results, discuss why you think this was done. Are there any potential problems with removing this variable?

Answer 4d)

- The above results shows that explanatory variable Pressure is statistically insignificant in both binomial and binary logistic model, when assessed by either Wald or Profile Likelihood statistics or confidence intervals. Regression table below is provided to compare each coefficient with its Wald standard error. In either model, point estimate for coefficient of pressure is close to its standard error, thus null hypothesis that it's insignificant cannot be rejected.

```r
# regression table
suppressWarnings(stargazer::stargazer(O.failed.binomial.glm, O.failed.binary.glm,
                    type = "text", title = "Binomial vs Binary Models",
                    star.cutoffs = c(0.05, 0.01, 0.001),
                    column.labels = c("binomial model", "binary model"),
                    model.names = F
                    ))
```

```
##
## Binomial vs Binary Models
## =====================================================
##                          Dependent variable:
```

```
##                       --------------------------------
##                       O.failed.binomial O.failed.binary
##                         binomial model    binary model
##                              (1)              (2)
## --------------------------------------------------------
## Temp                       -0.098           -0.229*
##                            (0.110)          (0.110)
##
## Pressure                    0.008            0.010
##                            (0.019)          (0.009)
##
## Constant                    2.520           13.292
##                            (8.541)          (7.664)
##
## --------------------------------------------------------
## Observations                 23               23
## Log Likelihood             -1.614           -9.391
## Akaike Inf. Crit.           9.229           24.782
## ========================================================
## Note:                    *p<0.05; **p<0.01; ***p<0.001
```

- The author has also followed up by a bootstrapping exercise which extract 90% confidence intervals for predicted O-ring failure probabilities from resamples of data. Although details on the procedure were not provided, we have conducted a similar exercise below for discussion. Our bootstrapping exercise resampled 1000 samples from our original data with replacement, each of 200 observations. 95% confidence intervals for expected number of failure versus a temperature range from 0 to 90F is plotted to compare the scenarios when pressure is 50psi against 200 psi.

```r
# get 500 resamples row ids of 23 observations
resamples_id = replicate(500, sample(23, size = 23, replace = T, prob = NULL))
# train 500 models
beta.int.list = c()
beta.T.list = c()
beta.S.list = c()
for (i in 1:500){
  resample = df[resamples_id[,i],]
  resample.glm = glm(formula = O.failed.binomial ~ Temp + Pressure ,
                             family = quasibinomial(link = logit), data = resample)
  beta.int.list = c(beta.int.list, as.numeric(resample.glm$coefficients[1]))
  beta.T.list = c(beta.T.list, as.numeric(resample.glm$coefficients[2]))
  beta.S.list = c(beta.S.list, as.numeric(resample.glm$coefficients[3]))}

# given temperature and pressure. get 95% conf
# loop through 500 models
pred.interval = function(beta.int.list, beta.T.list, beta.S.list, pressure, temp){
  expected.failures = c()
  for (i in 1:500){
    linear = beta.int.list[i] + beta.T.list[i]*temp + beta.S.list[i]*pressure
    expected.failure = exp(linear)/(1+exp(linear))*6
    expected.failures = c(expected.failures, expected.failure)}
  lower = as.numeric(quantile(expected.failures, 0.025, na.rm = T))
  upper = as.numeric(quantile(expected.failures, 0.975, na.rm = T))
  data.frame(temperature = temp, lower = lower, upper = upper)
}
```

```r
# Predict prob failure at pressure = 50psi, temp loop through 0 to 90F
pred.50.df = data.frame()
for (temp in seq(from = 0, to = 90, by = 0.5)) {
    pred.50 = pred.interval(beta.int.list, beta.T.list, beta.S.list, 50, temp)
    pred.50.df = rbind(pred.50.df, pred.50)
    }

# Predict prob failure at pressure = 200psi, temp loop through 0 to 90F
pred.200.df = data.frame()
for (temp in seq(from = 0, to = 90, by = 0.5)) {
    pred.200.df = rbind(pred.200.df, pred.interval(beta.int.list, beta.T.list, beta.S.list, 200, temp))}

psi.50 = 50
psi.200 = 200

# base predicted values
curve(expr = predict.glm(object = O.failed.binomial.glm,
        newdata = data.frame(Temp= x, Pressure = psi.50), type = "response")*6,
        col = "red", xlim = c(0,90), ylim = c(0,6),xlab = "Temperature in F",
        ylab = "Expected Number of Incidents", panel.first = grid(col = "gray", lty = "dotted"),
        main = "Binomial Logistic Model: Low vs High Pressure")

curve(expr = predict.glm(object = O.failed.binomial.glm,
                        newdata = data.frame(Temp= x, Pressure = psi.200), type = "response")*6,
                        col = "blue", xlim = c(0,90), add = T)

# Plot CI bands
points(x = pred.50.df$temperature,y= pred.50.df$lower, pch = ".", col = "red")
points(x = pred.50.df$temperature,y= pred.50.df$upper, pch = ".", col = "red")
points(x = pred.200.df$temperature,y= pred.200.df$lower, pch = ".", col = "blue")
points(x = pred.200.df$temperature,y= pred.200.df$upper, pch = ".", col = "blue")

# Legend
legend(x = 50, y = 6, legend = c("Pressure = 50psi","95% confidence interval",
                                "Pressure = 200psi", "95% confidence interval"),
        lty = c("solid", "dotted", "solid","dotted"), col = c("red","red","blue","blue"), bty = "n")
```
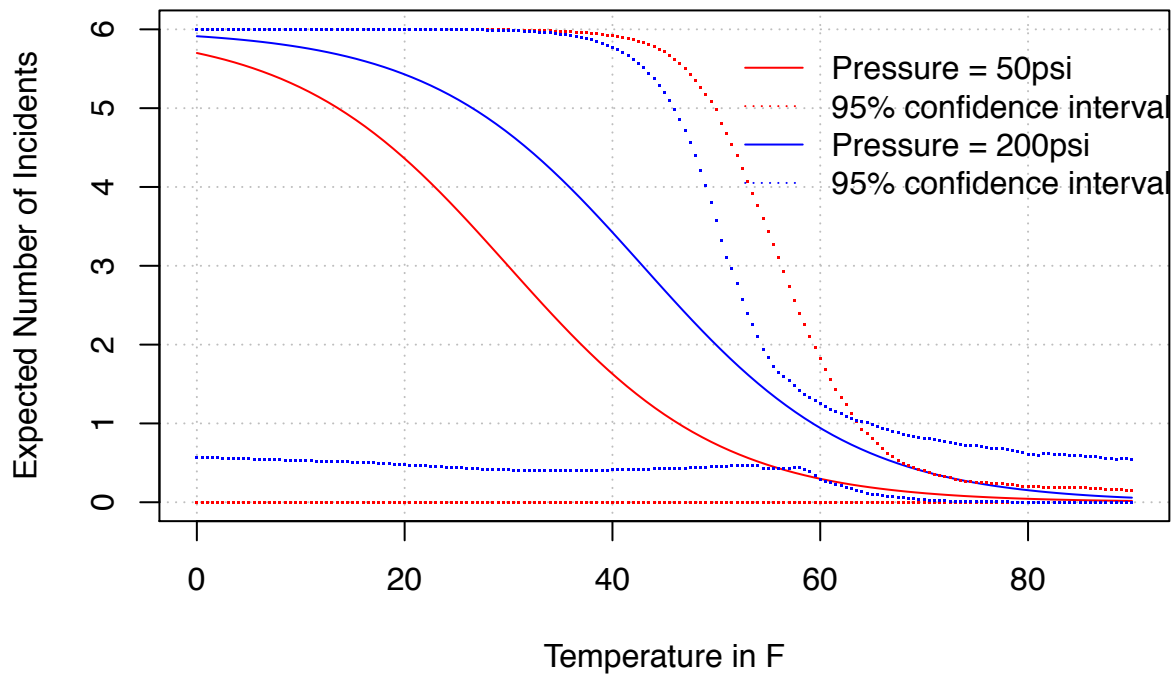
Binomial Logistic Model: Low vs High Pressure

- Notice that the two confidence interval bands overlap noticably, suggesting that the estimated number of O-ring failures is not significantly different at pressure 50psi and 200psi. At 31F and 200psi, the conditions under which the Challenger was last launched, our binomial logistic model predicts between 1.5 to 6 O-ring failures with a 95% confidence. If pressure was 50psi instead, the same model predicts 0 to 6 O-ring failures. Because of the extent to which these two intervals overlap, pressure's effect for this launching condition is not clear.

Binary Logistic Model: Low vs High Pressure
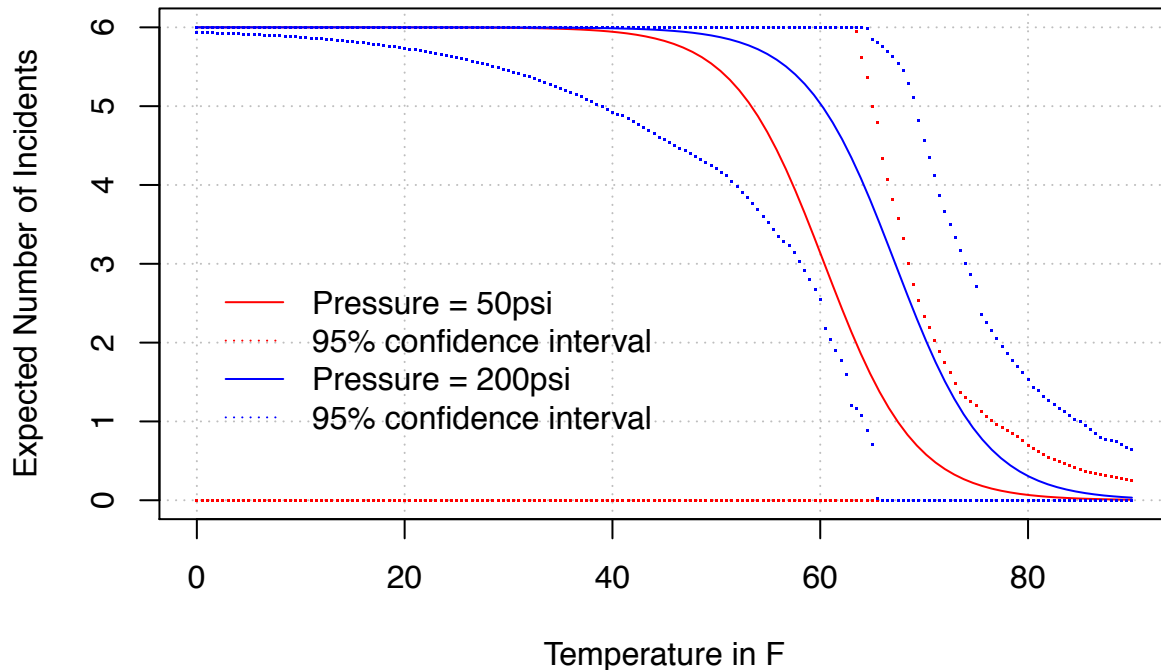
- We have conducted the same exercise for our binomial logistic regression model with the same explanatory variables and the same bootstrapped samples. A plot is provided above. The two confidence bands overlap less compared to the binomial model. However, this overlap persist for the full range of temperature. Once again we can't discriminate between the effects different pressure has on the dependent variable with 95% confidence.

- Thirdly, the author separated O-ring failure into O-ring erosion failure and O-ring blowby failure. Two separate logistic regression models are built and pressure has no effect on the response variable at all. Pressure was dropped as a variable after all three procedures.

- Although statistical evidence was not strong enough to support pressure as a plausible explanatory variable, domain knowledge and small sample should not be dismissed in the first place and thus we disagree with the author's decision.

  - First, the mechanics of a working O-ring depends on its elasticity, and the stress-strain curve in classic physics explains the relationship between stress(analogous to pressure) and strain(analogous to deformation) quite clearly. Once stress applied on material passes a certina yeild stress, O-ring material can loose its elasticity and deformation becomes permanent. The O-ring can either become plastic or fracture entirely, either of which can cause blowby and eventually O-ring failure.

  - Second, the plot in section 2.3.4 (EDA) of temperature and pressure illustrates that observations at 200psi and below 65F all experienced O-ring failure. We follow up here by training a binomial logistic model with temperature, pressure and interaction between the two. With interaction terms present, none of the estimated coefficients are statistically significant using likelihood ratio test, because the interaction term has reduced the unique variability in temperature and pressure themselves. From the likelihood ratio test p-value for the interaction term, we also know that the overall explanatory power of this model is not significantly different from its restricted model without any interactions.

- Ho: $\beta_{T:S} = 0$
- Ha: $\beta_{T:S} \neq 0$
- The test gave a p-value of 0.8264, therefore we don't have enough statistical evidence to reject the null hypothesis. But our sample size is too small and thus variability of our estimate is too large. A larger sample size and thus higher statistical power could have supported our intuition.

```
O.failed.binomial.interaction.glm = suppressWarnings(glm(formula = O.failed.binomial ~ Temp + Pressure
summary(O.failed.binomial.interaction.glm)
```

```
##
## Call:
## glm(formula = O.failed.binomial ~ Temp + Pressure + Temp:Pressure,
##     family = binomial(link = logit), data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.4197  -0.2694  -0.2099  -0.0712   0.9686
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -21.765547 113.376119  -0.192    0.848
## Temp            0.252463   1.623985   0.155    0.876
## Pressure        0.130857   0.569857   0.230    0.818
## Temp:Pressure  -0.001769   0.008172  -0.216    0.829
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4.0384  on 22  degrees of freedom
## Residual deviance: 2.7095  on 19  degrees of freedom
## AIC: 11.234
##
## Number of Fisher Scoring iterations: 7
```

```
suppressWarnings(car::Anova(O.failed.binomial.interaction.glm))
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: O.failed.binomial
##               LR Chisq Df Pr(>Chisq)
## Temp           0.86397  1     0.3526
## Pressure       0.25678  1     0.6123
## Temp:Pressure  0.04813  1     0.8264
```

```
anova(O.failed.binomial.glm,O.failed.binomial.interaction.glm,test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: O.failed.binomial ~ Temp + Pressure
## Model 2: O.failed.binomial ~ Temp + Pressure + Temp:Pressure
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        20     2.7576
## 2        19     2.7095  1 0.048126   0.8264
```

- A potential problem of dropping the variable pressure when it has an actual effect on O-ring failure is omitted variable bias on temperature. Applying intuitions from OLS regression, we know that if

pressure is omitted, we can estimating a biased coefficient for pressure, instead of its true population parameter.

$$\alpha_1 = \beta_1 + \beta_2 \delta_1$$

- where α_1 refers to the estimated coefficient of temperature in a model without variable pressure, β_1 refers to the true coefficient of temperature, β_2 refers to the true coefficient of pressure and δ_1 refers to the true coefficient if we regress pressure on temperature. $\beta_2 \delta_1$ is our omitted variable bias.

- From the intuition gained from our EDA, we assume increase in pressure increases the log odds (or odds) of O-ring failure thus $\beta_2 > 0$. Applying the Pressure-Temperature Law $P \propto T$ on the interiors of the solid rocket motors, we assume that gas temperature is directly proportional to gas pressure around each O-ring thus $\delta > 0$. Note that this is also agreed by our EDA where correlation between "Temp" and "Pressure" was found to be weakly, positively correlated. Therefore, our omitted variable bias $\beta_2 \delta_1$ is positive. If pressure had a real effect on log-odds of failure but we omit it, we would have over-estimated the value of coefficient of temperature.

Question 5

5a) Estimate the simplified model

- Below we are again providing two logistic regression models for binomial and binary response with the specification.

Answer 5a) 1.Binomial Logistic Regression Model

$$logit(\pi) = \beta_0 + \beta_T temp$$

```
O.failed.binomial_temp.glm = suppressWarnings(glm(formula = O.failed.binomial ~ Temp,
                          family = binomial(link = logit), data = df))

summary(O.failed.binomial_temp.glm)
```

```
##
## Call:
## glm(formula = O.failed.binomial ~ Temp, family = binomial(link = logit),
##     data = df)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.38876  -0.31965  -0.22093  -0.01788   1.08248
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   5.0850     7.4770   0.680    0.496
## Temp         -0.1156     0.1152  -1.004    0.316
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4.0384  on 22  degrees of freedom
## Residual deviance: 3.0144  on 21  degrees of freedom
## AIC: 7.2026
##
## Number of Fisher Scoring iterations: 6
```

```r
logLik(O.failed.binomial_temp.glm)
```

```
## 'log Lik.' -1.601316 (df=2)
```

- Point estimations of the coefficients in our binomial model agree with that of the author's.

- Temperature (-0.1156) again has a negative sign, suggesting an increase of temperature decreases expected odds of O-ring failure. The estimated odds of O-ring failure changes by 0.442 times, equivalent to decrease by around 55.8%, with one standard deviation of increase (7.057F) in temperature.

$$OR = \frac{Odds_{T+c}}{Odds_T} = \frac{e^{\beta_0 + \beta_T(T+c)}}{e^{\beta_0 + \beta_T T}} = e^{c\beta_T} = e^{7.057*(-0.116)} = 0.442$$

- Regarding the significance of our estimates, Wald Test uses the hypotheses:

  - Ho: the estimated coefficient is zero.
  - Ha: the estimated coefficient is not zero.

$$Z_o = \frac{\hat{\beta}_r - 0}{\sqrt{Var(\hat{beta}_r)}}$$

- From the default regression output, with a 95% confidence, coefficient for temperature is statistically insignificant (p-value) using Wald statistics. Wald Confidence Interval covers zero and agrees that the null hypothesis cannot be rejected.

- If we "invert" the test statistic, we get Wald Confidence Interval:

$$\hat{\pi} \pm Z_{1-\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

where coefficients are linked to probability estimates by:

$$\hat{\pi} = \frac{e^{\beta_0 + \beta_t Temperature}}{1 + e^{\beta_0 + \beta_t Temperature}}$$

```r
confint.default(object = O.failed.binomial_temp.glm, level = 0.95)
```

```
##                   2.5 %     97.5 %
## (Intercept) -9.5697302 19.739685
## Temp        -0.3413584  0.110156
```

- Below, we use the more reliable Likelihood Ratio Test to assess significance of the coefficients. The restricted model in our case only estimate the coefficient for the intercept.

  - Ho: $\beta_r = 0$

  - Ha: $\beta_r \neq 0$

  - With a p-value of 0.312 and the Profile Likelihood Ratio Interval covers zero, we fail to reject the null hypothesis. The Likelihood Ratio Test results agrees with Wald Test results.

```r
suppressWarnings(car::Anova(O.failed.binomial_temp.glm))
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: O.failed.binomial
##      LR Chisq Df Pr(>Chisq)
## Temp    1.024  1     0.3116
```

```
PLR.binomial =suppressWarnings(confint(object = O.failed.binomial_temp.glm, level = 0.95))
PLR.binomial
```

```
##                   2.5 %     97.5 %
## (Intercept) -11.1396429 21.3839149
## Temp          -0.3830992  0.1172154
```

```
par(mfrow = c(2, 2))
plot(O.failed.binomial_temp.glm)
```
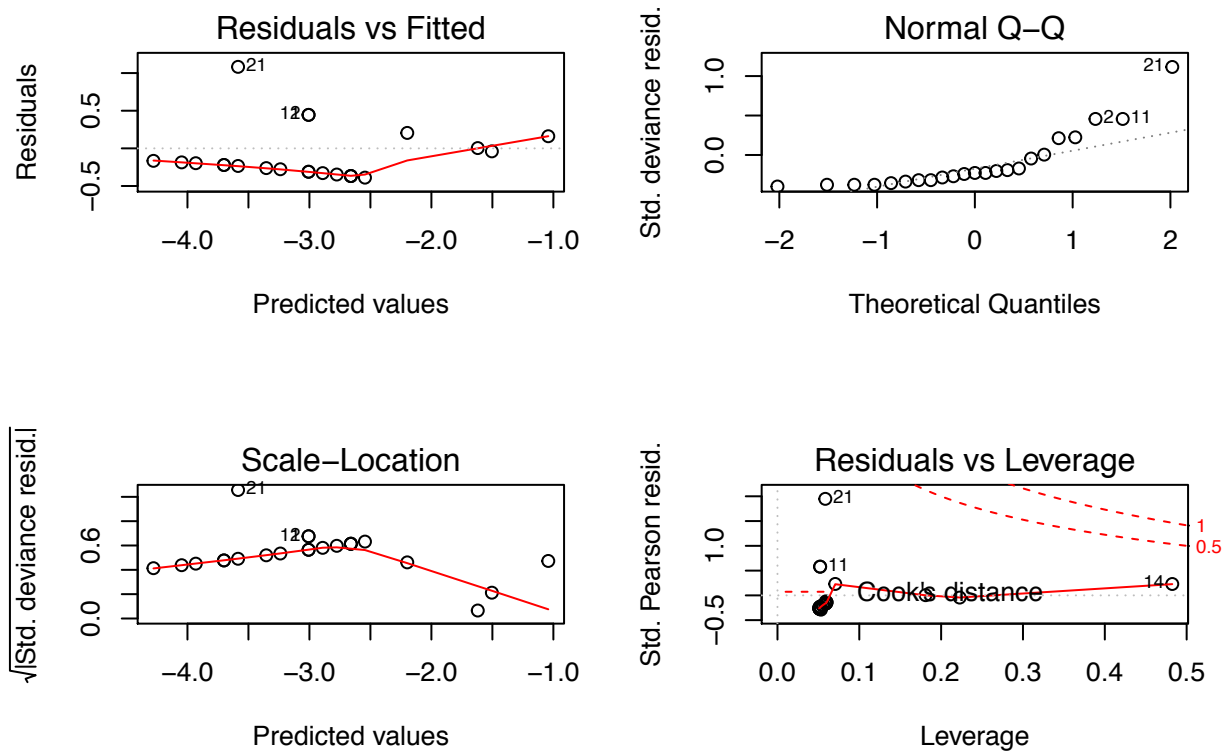

- From the plots above, we compare the residual behaviors against the binomial model without "Pressure" and highlight the differences:

  - Zero-conditional mean violation: the mean of our residuals have a dip in the middle at fitted values of -2.5. The residuals' distribution is similar but without a persistent downward trend.

  - Heteroskedasticity: Residuals' trend is similar but less dispersed to the top. Breusch Pagan Test again fail rejected the null hypothesis of homoskedasticity. Due to small sample size robust Standard Erros are recommended and provided below.

  - H_o : variance of our residual changes with fitted values

  - H_a : variance of our residual does not change with fitted values

```
lmtest::bptest(O.failed.binomial_temp.glm)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  O.failed.binomial_temp.glm
```

```
## BP = 0.04457, df = 1, p-value = 0.8328
# White standard errors
# lmtest::coeftest(O.failed.binomial_temp.glm, vcov = vcovHC)
sqrt(diag(vcovHC(O.failed.binomial_temp.glm)))

## (Intercept)        Temp
##  3.24047127  0.05363397
```

- Outliers: Only observation 21 has a standardized Pearson's residual close to 2. Other outliers such as observation 2 and 11 are much more tamed now, with value less than 1.

- We take note of the residual deviance (3.0144), log likelihood (-1.601) and AIC (7.2026). The model has performed better in terms of residual deviance and AIC than the model with "Pressure".

- From the Likelihood ratio test performed in section 4c), we already know that the model including "Pressure" as an explanatory variable doesn't have significantly higher overall explanatory power than the restricted model (p value 0.6123). So we skip the test here.

5a) 2.Binary Logistic Regression Model

```
O.failed.binary_temp.glm = glm(formula = O.failed.binary ~ Temp,
                               family = binomial(link = logit), data = df)

summary(O.failed.binary_temp.glm)

##
## Call:
## glm(formula = O.failed.binary ~ Temp, family = binomial(link = logit),
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0611  -0.7613  -0.3783   0.4524   2.2175
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  15.0429     7.3786   2.039   0.0415 *
## Temp         -0.2322     0.1082  -2.145   0.0320 *
## ---
## 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: 20.315  on 21  degrees of freedom
## AIC: 24.315
##
## Number of Fisher Scoring iterations: 5
```

```
logLik(O.failed.binary_temp.glm)

## 'log Lik.' -10.1576 (df=2)
```

- Point estimations of the coefficients in our binary model agree with that of the author's.

- Temperature (-0.2322) again has a negative sign, suggesting an increase of temperature decreases expected odds of O-ring failure. The estimated odds of O-ring failure changes by 0.194 times, equivalent to decrease by around 81%, with one standard deviation of increase (7.057F) in temperature.

$$OR = \frac{Odds_{T+c}}{Odds_T} = \frac{e^{\beta_0 + \beta_T(T+c)}}{e^{\beta_0 + \beta_T T}} = e^{c\beta_T} = e^{7.057*(-0.2322)} = 0.194$$

```
confint.default(object = O.failed.binary_temp.glm, level = 0.95)
```

```
##                   2.5 %      97.5 %
## (Intercept)   0.5810523 29.50475096
## Temp         -0.4443022 -0.02002324
```

- Point estimations of the coefficients in our binomial model agree with that of the author's.

- From the default regression output, with a 95% confidence, coefficient for temperature is statistically significant (p-value) using Wald statistics. Wald Confidence Interval clears zero. The result is quite close to the binary model without "Pressure" as explanatory variable/

- Below, we use the more reliable Likelihood Ratio Test to assess significance of the coefficients. The restricted model in our case only estimate the coefficient for the intercept.

  - Ho: $\beta_r = 0$

  - Ha: $\beta_r \neq 0$

  - With a p-value of 0.0048 and the Profile Likelihood Ratio Interval clears zero, the null hypothesis can be rejected. The Likelihood Ratio Test results agrees with Wald Test results.

```
car::Anova(O.failed.binary_temp.glm)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: O.failed.binary
##       LR Chisq Df Pr(>Chisq)
## Temp    7.952  1   0.004804 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
PLR.binary = confint(object = O.failed.binary_temp.glm, level = 0.95)
PLR.binary
```
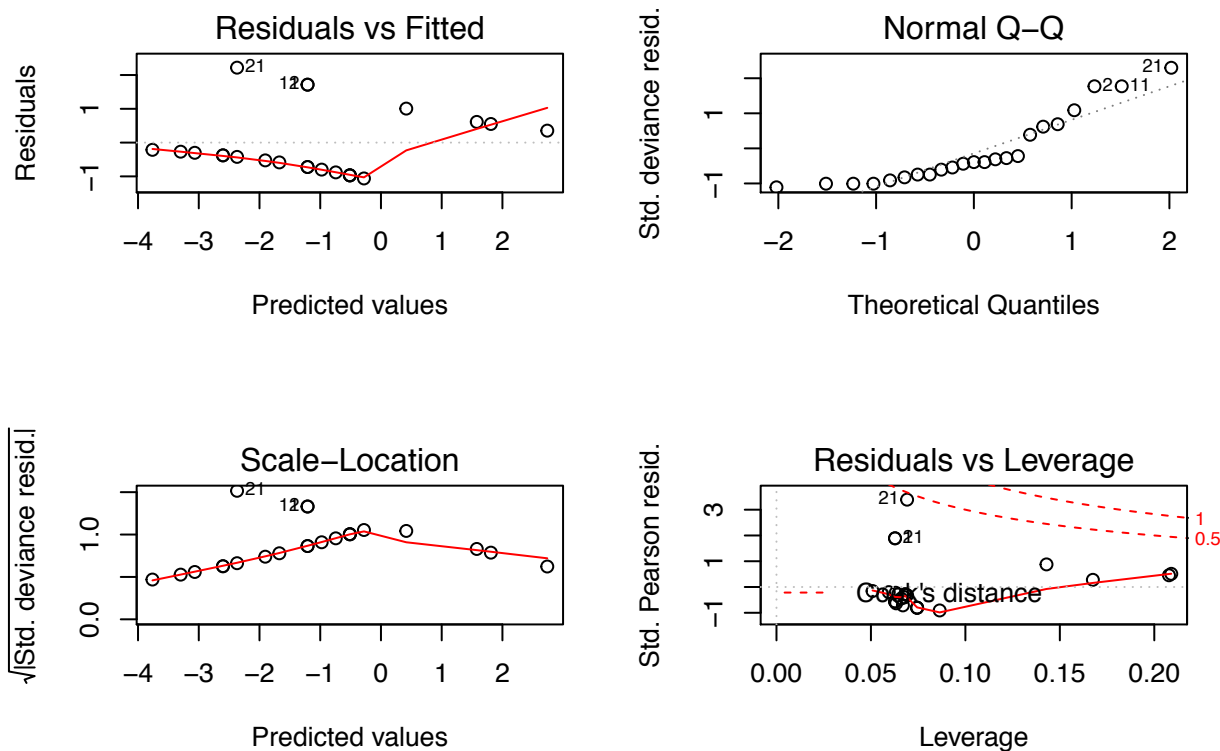
```
##                   2.5 %      97.5 %
## (Intercept)   3.3305848 34.34215133
## Temp         -0.5154718 -0.06082076
```

```
par(mfrow = c(2, 2))
plot(O.failed.binary_temp.glm)
```

- The residuals plots compared with the binary logistic model that includes "Pressure" is almost identical. The only key difference is the extremity of outliers. In the current mode, observation 21 is behaving more extreme than observation 11 which is behaving similar to observation 2. In the model which "Pressure" is included, observation 2 is behaving more extreme than observation 21 which is behaving more extreme than observation 11. Closer examination of these data points reveals that the three observations had similar temperature and all experienced one or more O-ring failures. But observation 2 has a pressure of 50psi, that is 150 psi lower than the other two. That may explain why the two models rank their extremity differently.

```
df[c(21,11,2),]
```

```
##    Flight Temp Pressure O.ring Number O.failed.binary O.failed.binomial
## 21     21   75      200      2      6               1         0.3333333
## 11     11   70      200      1      6               1         0.1666667
## 2       2   70       50      1      6               1         0.1666667
```

- We take note of the residual deviance (20.315), log likelihood (-10.1576) and AIC (24.315). The model has performed worse in terms of residual deviance (expected with removal of variable "Pressure") but slight better in terms of AIC than the binary model with "Pressure" present.

5b) Construct two plot: (1) pi vs. Temp (2) Expected number of failures vs Temp. Use the temp range of 31 to 81 on the x-axis even though the minimum temp in the data set was 53

Answer 5b, c)

```r
# function to estimate CI for input distance = x
ci.pi = function(newdata, mod.fit.obj, alpha) {
  linear.pred = predict.glm(object = mod.fit.obj, se = TRUE, type = "link", newdata = newdata)
  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)
}

# base plot of observations
plot(x = df$Temp, y=jitter(df$O.ring,0.2), xlab = "Temperature in F",
     ylab = "Expected Number of Incidents",
     panel.first = grid(col = "gray", lty = "dotted"), ylim = c(0,6), xlim = c(30,81),
     main = "Expected number of O-ring failures vs Temperature ")

curve(expr = predict(object = O.failed.binomial_temp.glm,
      newdata = data.frame(Temp= x), type = "response")*6,
      col = "red", xlim = c(31,81), add = T)

curve(expr = predict(object = O.failed.binary_temp.glm,
      newdata = data.frame(Temp= x), type = "response")*6,
      col = "blue", xlim = c(31,81), add = T)

# Plot CI bands
curve(expr = ci.pi(newdata = data.frame(Temp = x), mod.fit.obj = O.failed.binomial_temp.glm, alpha = 0.0
      xlim = c(31,81), add = T, lty = "dotted", col = "red")
curve(expr = ci.pi(newdata = data.frame(Temp = x), mod.fit.obj = O.failed.binomial_temp.glm, alpha = 0.0
      xlim = c(31,81), add = T, lty = "dotted", col = "red")

curve(expr = ci.pi(newdata = data.frame(Temp = x), mod.fit.obj = O.failed.binary_temp.glm, alpha = 0.05)
      xlim = c(31,81), add = T, lty = "dotted", col = "blue")
curve(expr = ci.pi(newdata = data.frame(Temp = x), mod.fit.obj = O.failed.binary_temp.glm, alpha = 0.05)
      xlim = c(31,81), add = T, lty = "dotted", col = "blue")

# Legend
legend(x = 28, y = 2.5, legend = c("Binomial Logit model", "95% Confidence Interval","Binary Logit mode
       lty = c("solid","dotted", "solid", "dotted"), col = c("red","red","blue", "blue"), bty = "n")
legend(x = 31, y = 3, legend = c("observations"),
       pch = 1, bty = "n")
```
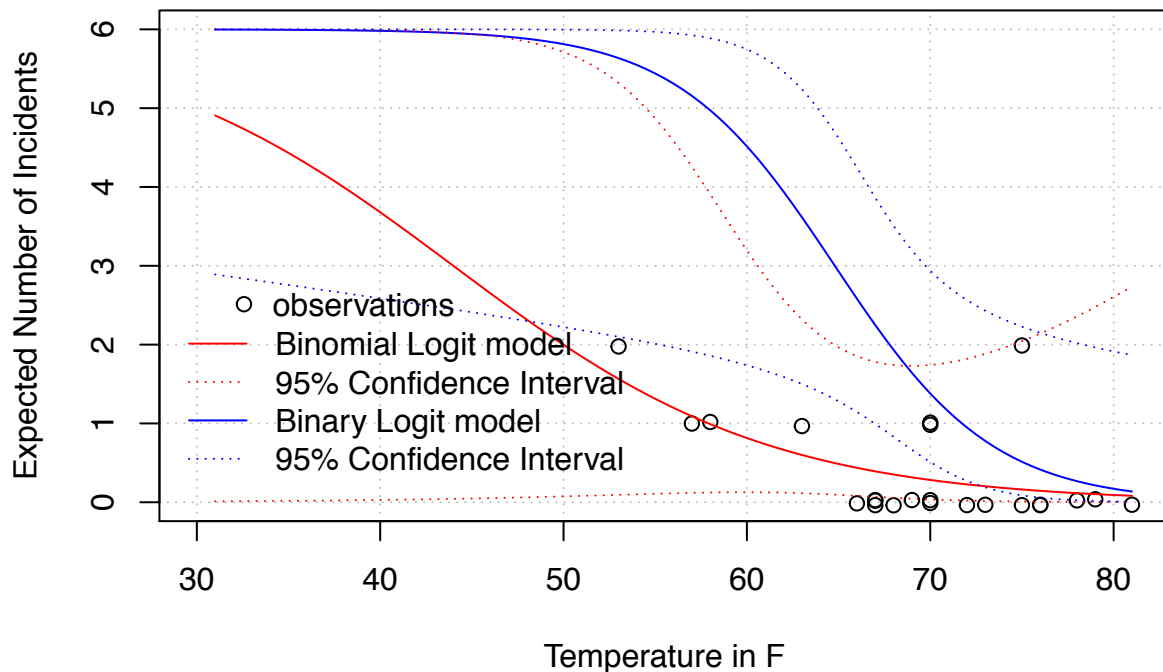
Expected number of O–ring failures vs Temperature

```r
# base plot of observations
y = df$O.ring/6

plot(x = df$Temp, y = jitter(y,0.02), xlab = "Temperature in F",
     ylab = "Probability of O-ring failure",
     panel.first = grid(col = "gray", lty = "dotted"), ylim = c(0,1), xlim = c(30,81),
     main = "Predicted Probability of O-ring failure vs Temperature ")

curve(expr = predict(object = O.failed.binomial_temp.glm,
      newdata = data.frame(Temp= x), type = "response"),
      col = "red", xlim = c(31,81), add = T)

curve(expr = predict(object = O.failed.binary_temp.glm,
      newdata = data.frame(Temp= x), type = "response"),
      col = "blue", xlim = c(31,81), add = T)

# Plot CI bands
curve(expr = ci.pi(newdata = data.frame(Temp = x), mod.fit.obj = O.failed.binomial_temp.glm, alpha = 0.0
      xlim = c(31,81), add = T, lty = "dotted", col = "red")
curve(expr = ci.pi(newdata = data.frame(Temp = x), mod.fit.obj = O.failed.binomial_temp.glm, alpha = 0.0
      xlim = c(31,81), add = T, lty = "dotted", col = "red")

curve(expr = ci.pi(newdata = data.frame(Temp = x), mod.fit.obj = O.failed.binary_temp.glm, alpha = 0.05]
      xlim = c(31,81), add = T, lty = "dotted", col = "blue")
curve(expr = ci.pi(newdata = data.frame(Temp = x), mod.fit.obj = O.failed.binary_temp.glm, alpha = 0.05]
      xlim = c(31,81), add = T, lty = "dotted", col = "blue")
```
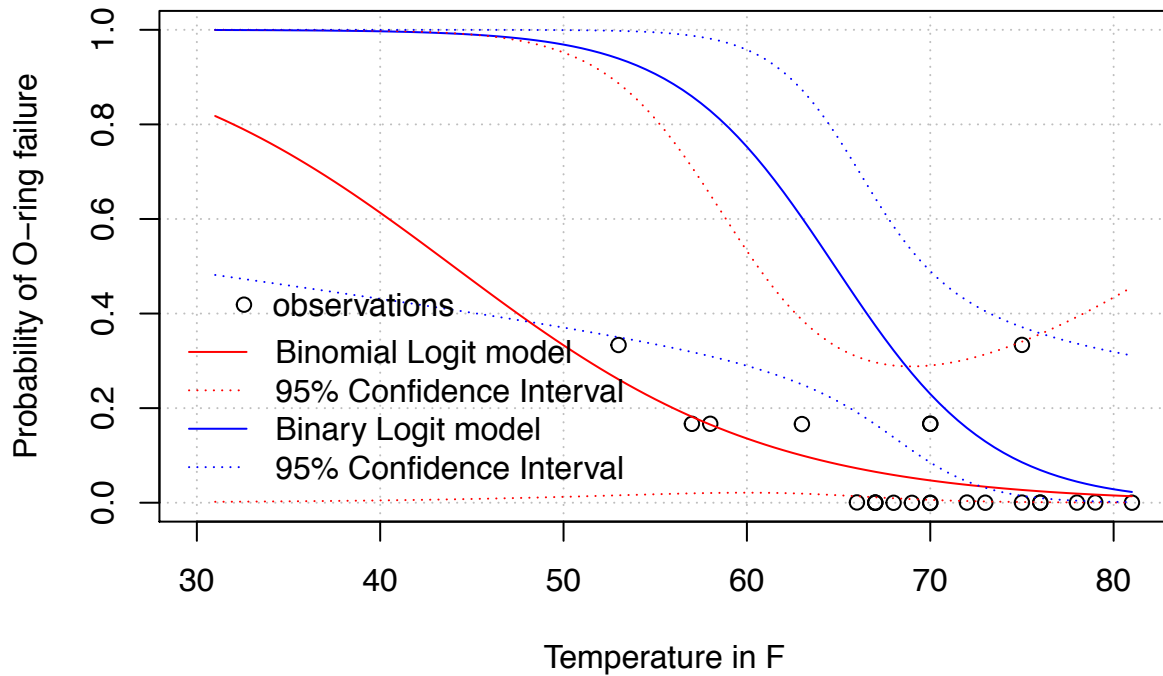
```
# Legend
legend(x = 28, y = 0.4, legend = c("Binomial Logit model", "95% Confidence Interval","Binary Logit model
    lty = c("solid","dotted", "solid", "dotted"), col = c("red","red","blue", "blue"), bty = "n")
legend(x = 31, y = 0.5, legend = c("observations"),
    pch = 1, bty = "n")
```

Predicted Probability of O–ring failure vs Temperature

- Similar to the comparison we have done for the models which included "Pressure" as an explanatory variable, the binary and binomial models here agrees that decrease in temperature is estimated to increase probability of O-ring failure. The binary model has a steeper fitted curve and expect between 4 to 6 O-rings to fail once temperature drops below 40F. The binomial model has a smoother curve and expects between 0 to 6 O-rings to fail once temperature drops below 40F. The binomial model is not very informative to estimate probability of failure at the launching condition of 31F, the temperature at which our study is most interested in.

- The confidence bands are much wider for lower temperature because we have very few observations in that region. The Wald confidence interval defines a region around the expected value that is proportional to variance of the predicted value. The variance of this predicted value dependes on sample size n. Once n drops, that is where we have fewer observations, this variance goes up and our confidence band becomes wider. Most of our observations are gathered between 60F to 80F, thus the confidence band is wider in that region.

$$\hat{\pi} \pm Z_{1-\alpha/2}\sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n}}$$

5d) The temperature was 31F 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.

Answer 5d)

- Below we provide estimations of the probability of O-ring failure under 31F, temperature at which the Challenger was launched. Corresponding Wald and Profile Likelihood Ratio Confidence Intervals are also provided for comparison between the binary and binomial logistic regression model.

```r
# Prediction and Wald CI for binomial model
lin.pred.binomial = predict(O.failed.binomial_temp.glm,newdata= data.frame(Temp = 31),type = 'link', se
pi.hat.binomial = exp(lin.pred.binomial$fit) / (1 + exp(lin.pred.binomial$fit))
wald.ci.pi.binomial = ci.pi(data.frame(Temp = 31),O.failed.binomial_temp.glm,0.05 )
wald.ci.width.binomial = wald.ci.pi.binomial$upper - wald.ci.pi.binomial$lower
# Prediction and Wald CI for binary model
lin.pred.binary = predict(O.failed.binary_temp.glm,newdata= data.frame(Temp = 31),type = 'link', se = T
pi.hat.binary = exp(lin.pred.binary$fit) / (1 + exp(lin.pred.binary$fit))
wald.ci.pi.binary = ci.pi(data.frame(Temp = 31),O.failed.binary_temp.glm,0.05 )
wald.ci.width.binary = wald.ci.pi.binary$upper - wald.ci.pi.binary$lower


# Profile LR Interval for binomial model
K = matrix(data = c(1,31), nrow = 1, ncol = 2)
# calculate -2log(Lambda)
linear.combo.binomial = suppressWarnings(mcprofile(object = O.failed.binomial_temp.glm, CM = K))
# CI for linear combination
profile.ci.lincombo.binomial = suppressWarnings(confint(object = linear.combo.binomial, level = 0.95))
profile.ci.binomial = exp(profile.ci.lincombo.binomial$confint)/(1 + exp(profile.ci.lincombo.binomial$co
profile.ci.width.binomial = profile.ci.binomial$upper - profile.ci.binomial$lower


# Profile LR Interval for binary model
# calculate -2log(Lambda)
linear.combo.binary = suppressWarnings(mcprofile(object = O.failed.binary_temp.glm, CM = K))
# CI for linear combination
profile.ci.lincombo.binary = suppressWarnings(confint(object = linear.combo.binary, level = 0.95))
profile.ci.binary = exp(profile.ci.lincombo.binary$confint)/(1 + exp(profile.ci.lincombo.binary$confint)
profile.ci.width.binary = profile.ci.binary$upper - profile.ci.binary$lower

data.frame(model = c("Binomial Logistic Model", "Binary Logistic Model"),
           pi.hat = c(pi.hat.binomial, pi.hat.binary),
           Wald.ci.lower = c(wald.ci.pi.binomial$lower, wald.ci.pi.binary$lower),
           Wald.ci.upper = c(wald.ci.pi.binomial$upper, wald.ci.pi.binary$upper),
           Wald.ci.width = c(wald.ci.width.binomial, wald.ci.width.binary),
           Profile.LR.ci.lower = c(profile.ci.binomial$lower, profile.ci.binary$lower),
           Profile.LR.ci.upper = c(profile.ci.binomial$upper, profile.ci.binary$upper),
           Profile.LR.ci.width = c(profile.ci.width.binomial, profile.ci.width.binary))
```

```
##                     model    pi.hat Wald.ci.lower Wald.ci.upper
## 1 Binomial Logistic Model 0.8177744   0.001935866     0.9999037
## 2   Binary Logistic Model 0.9996088   0.481610583     0.9999999
##   Wald.ci.width Profile.LR.ci.lower Profile.LR.ci.upper
## 1     0.9979678        0.0005181288           0.999935
## 2     0.5183893        0.8036982445           1.000000
##   Profile.LR.ci.width
## 1           0.9994169
```

```
## 2            0.1963017
```

- The binary model predicted a higher probability of O-ring failure at 0.9996 than the binomial model at 0.8178. We also notice that the confidence interval, either Wald or Profile Likelihood Ratio, are much narrower. The binomial model produced confidence intervals, either Wald or Profile Likelihood Ratio, of width close to 1 so the estimation is indeed not very useful.

- To derive Wald"s confidence interval, there needs to be a large sample following a distribution of the exponential family so the sampling distribution will converge to a normal distribution approximately and asymptotically. The confidence interval is essential an "inverse" of the Wald's Test statistic and it has the form cited at the end of section 5c). Given our small sample size, we can't realistically rely on it. In addition, the true confidence level of the Wald's confidence interval fluctuates drastically when predicted probability is close to zero or one. This is true for our predicted probability of 0.8178 in the binomial model and 0.9996 in the binary model. Therefore, for the binary model predictions, we see that the Wald confidence interval of width 0.5184 is noticably wider than the profile likelihood ratio interval of width 0.1963.

- To derive the Profile Likelihood Ratio interval, besides having an underlying binomial distribution of our response variable, our observations also needs to be independently generated across all records. If a response variable is a binomial response, then each of its trials – the Bernoulli responses, need to be independet as well. This is because calculation of the Likelihood Ratio Interval rely on the calculation of a Chi-square test statistic that is derived from maximum likelihood functions. We have already discussed more details of this requirement section 4a). Our binomial model has obivously violated this assuptiom, so we should rely on the binary model's estimates instead.

$$-2[log(L(\tilde{\beta}_0, \beta_T|y)) - log(L(\hat{\beta}_0, \hat{\beta_T}|y))] - \chi^2_{1,0.95} < 0$$

where:

$$L(\beta_0, ..., \beta_p|y_1, ..y_n) = \prod_{i=1}^{n} \pi_i^{y_i}(1 - \pi_i)^{(1-y_i)}$$

5e) Using parametric bootstrapping, compute the 90% confidence intervals separately at temperatures of 31 and 72.

```
nBoot = 1000

# Bootstrap a large dataset each with 23 temp measurements
large.dataset = replicate(nBoot, sample_n(data.frame(df$Temp,df$O.failed.binomial), 23, replace=TRUE))

# Update Binomial by simulating the probability from sampled temperature
for (i in 1:nBoot){
  dfi = data.frame(prob = large.dataset[,i]$df.O.failed.binomial, Temp = large.dataset[,i]$df.Temp)
  model = suppressWarnings(glm(formula  = prob ~ Temp, family = quasibinomial(link = "logit"), data = d:
  updateProb = predict(object = model, data.frame(Temp = dfi$Temp),type='response')
  # Update
  large.dataset[,i]$df.O.failed.binomial = updateProb
}

# Refit the model and get 90% confidence level for temperature at temptoExamine
bootstrap90 <- function(temptoExamine, large.dataset, nBoot){
  probability.failure = c()
  for (i in 1:nBoot){
    dfi = data.frame(prob = large.dataset[,i]$df.O.failed.binomial, Temp = large.dataset[,i]$df.Temp)
```

```
    model = suppressWarnings(glm(formula  = prob ~ Temp, family = quasibinomial(link = "logit"), data =


    logit = model$coefficient[1] + model$coefficient[2]*temptoExamine
    bootstrap.prob = exp(logit)/(1+exp(logit))
    probability.failure = c(probability.failure,bootstrap.prob)
  }

  lower= quantile(probability.failure,0.05)
  upper= quantile(probability.failure,0.95)
  list(lower,upper)

}


print("90% CI at 31 F is")
```

```
## [1] "90% CI at 31 F is"
```

```
bootstrap90(31,large.dataset,nBoot)
```

```
## [[1]]
##        5%
## 0.1257348
##
## [[2]]
##      95%
## 0.98943
```

```
print("90% CI at 72 F:")
```

```
## [1] "90% CI at 72 F:"
```

```
bootstrap90(72,large.dataset,nBoot)
```

```
## [[1]]
##          5%
## 0.006760624
##
## [[2]]
##        95%
## 0.0751623
```

- We generate 1000 (n=23) data sets from the the original data sets. From these data sets, we estimate the first sets of parameters, which we use to get 1000 (n=23) simulated data sets. By feeding the simulated data sets back into logistic (binomial) regression, we obtain new estimates for the parameters and their probabilities of failures.

- Bootstrap provides a tighter confidence intervals than the wald method. This is based roughly on the law of large numbers or the assumption that with enough data we can build an empirical distribution that can b~~e a good approximation of the true distribution to~~ approximate the true distribution.

5f) Determine if a quadratic term is needed in the model for the temperature.

Answer 5f)

- From our univariate analysis of "Temp" in section 2.2.2, we know that a square or cubic term of "Temp" can reduce negative skew of the variable. We have also seen from residual plots of previous models about systematic violations of zero-conditional mean and heteroskedasticity of obvious residual trends. A high order term may be necessary for the model and in this section we construct new models to test this.

```
par(mfrow = c(2, 2))
plot(O.failed.binomial.glm, which = 1, caption = "Binomial Logit: Temperature + Pressure" )
plot(O.failed.binary.glm, which = 1, caption = "Binary Logit: Temperature + Pressure")
plot(O.failed.binomial_temp.glm, which = 1, caption = "Binomial Logit: Temperature")
plot(O.failed.binary_temp.glm, which = 1, caption = "Binary Logit: Temperature")
```
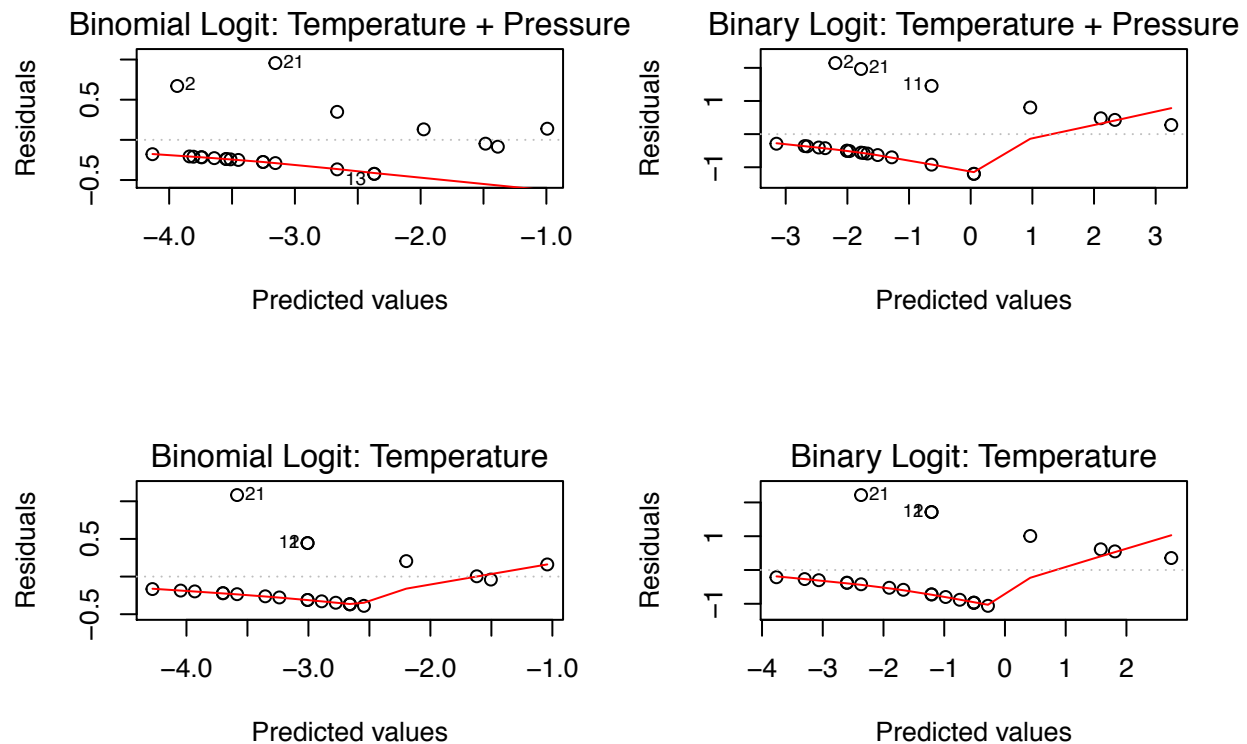

- The correlation value between "Temp" and "O.ring" grow slightly closer to zero as we impose higher oder transformation on temperature.

```
O.failed.binomial_temp_sq.glm = suppressWarnings(glm(formula = O.failed.binomial ~ Temp + I(Temp^2),
                        family = binomial(link = logit), data = df))

O.failed.binary_temp_sq.glm = suppressWarnings(glm(formula = O.failed.binary ~ Temp + I(Temp^2),
                        family = binomial(link = logit), data = df))

O.failed.binomial_temp_sq_cu.glm = suppressWarnings(glm(formula = O.failed.binomial ~ Temp + I(Temp^2) +
                        family = binomial(link = logit), data = df))
```

```
O.failed.binary_temp_sq_cu.glm = suppressWarnings(glm(formula = O.failed.binary ~ Temp + I(Temp^2) + I(
                            family = binomial(link = logit), data = df))
```

- In the models with the square term added, none of the coefficients are significant. Specifically in the binary model, temperature has become statistically insignificant with the addition of a squared-term. This addition have introduced enough correlation between explanatory variables that variance of our coefficient estimates has increased, and their confidence intervals has widened enough to cover zero. Therefore, we can no longer reject the null hypothesis that coefficient of "Temp" is not zero.

```
summary(O.failed.binomial_temp_sq.glm)
```

```
##
## Call:
## glm(formula = O.failed.binomial ~ Temp + I(Temp^2), family = binomial(link = logit),
##     data = df)
##
## Deviance Residuals:
##       Min         1Q     Median         3Q        Max
## -0.34423   -0.29551   -0.25303   -0.00545    1.02920
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) 22.126148  58.284482    0.380    0.704
## Temp        -0.650885   1.814484   -0.359    0.720
## I(Temp^2)    0.004141   0.013943    0.297    0.766
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4.0384  on 22  degrees of freedom
## Residual deviance: 2.9319  on 20  degrees of freedom
## AIC: 9.2373
##
## Number of Fisher Scoring iterations: 6
```

```
summary(O.failed.binary_temp_sq.glm)
```

```
##
## Call:
## glm(formula = O.failed.binary ~ Temp + I(Temp^2), family = binomial(link = logit),
##     data = df)
##
## Deviance Residuals:
##     Min       1Q    Median       3Q       Max
## -0.9614   -0.6608   -0.4735    0.1781    2.1185
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 78.48317   76.67353    1.024    0.306
## Temp        -2.09430    2.18479   -0.959    0.338
## I(Temp^2)    0.01359    0.01553    0.875    0.381
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 28.267  on 22  degrees of freedom
## Residual deviance: 19.389  on 20  degrees of freedom
```

```
## AIC: 25.389
##
## Number of Fisher Scoring iterations: 6
```

- Likelihood Ratio Test also agree with the Wald Test results. With p-values well over 0.05, we again failed to reject the null hypothesis that the full and restricted model has the same overall explanatory power. Result is similar for both the binary and the binomial models.

  - $\beta_T^2 = 0$
  - $\beta_T^2 \neq 0$

```
Anova(O.failed.binomial_temp_sq.glm)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: O.failed.binomial
##           LR Chisq Df Pr(>Chisq)
## Temp       0.11980  1     0.7293
## I(Temp^2)  0.08245  1     0.7740
```

```
Anova(O.failed.binary_temp_sq.glm)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: O.failed.binary
##           LR Chisq Df Pr(>Chisq)
## Temp       1.19322  1     0.2747
## I(Temp^2)  0.92649  1     0.3358
```

- With the cubic term as an addition, p values of all other coefficients have gone up further more in both models.

```
Anova(O.failed.binomial_temp_sq_cu.glm)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: O.failed.binomial
##           LR Chisq Df Pr(>Chisq)
## Temp      0.074089  1     0.7855
## I(Temp^2) 0.066457  1     0.7966
## I(Temp^3) 0.061162  1     0.8047
```

```
Anova(O.failed.binomial_temp_sq_cu.glm)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: O.failed.binomial
##           LR Chisq Df Pr(>Chisq)
## Temp      0.074089  1     0.7855
## I(Temp^2) 0.066457  1     0.7966
## I(Temp^3) 0.061162  1     0.8047
```

- Below we compare residuals vs fitted value plots of the models with square term to the restricted models. We notice that the models with square terms violate zero-conditional mean with a clear downward residual trend, while the resticted models takes off and return back to reference line of zero. Therefore, the violation of our square term models seem stronger. The addition of higher order terms for temperature increased the variance of our estimates without improving bias of the model. There isn't a bias variance trade-off. We conclude that a higher order term for temperature is not neeeded.
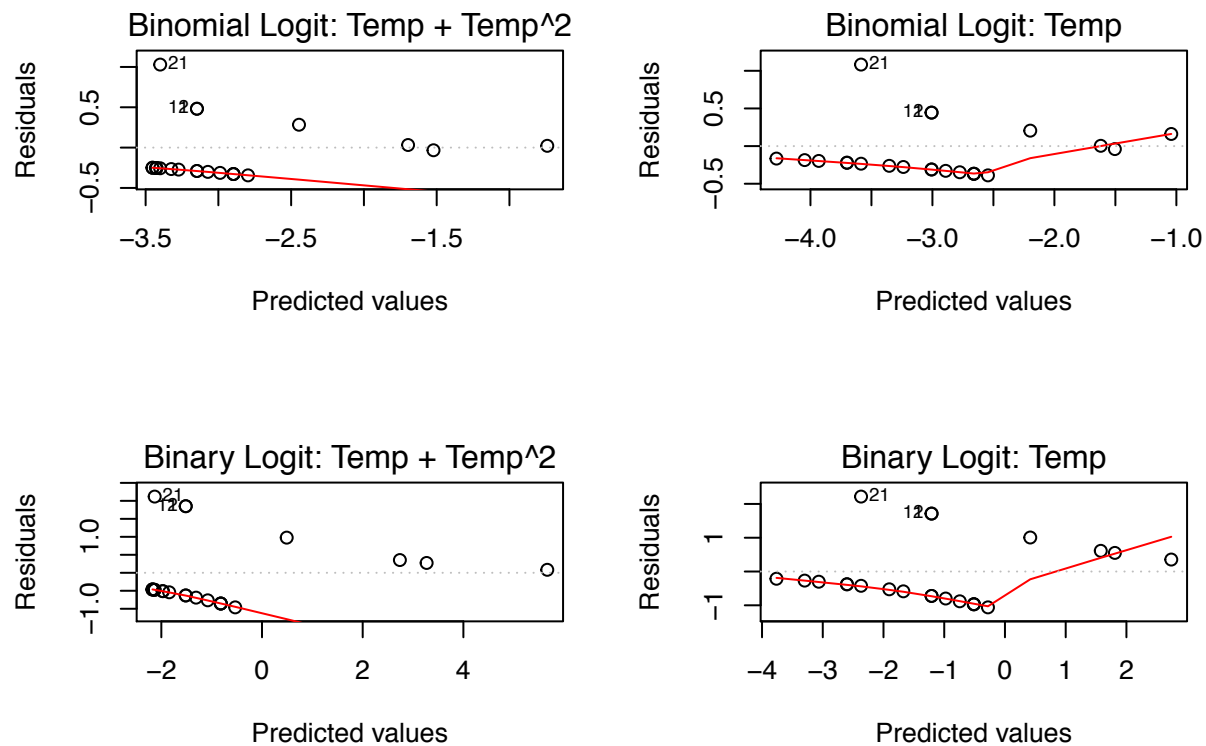
```
par(mfrow = c(2, 2))
plot(O.failed.binomial_temp_sq.glm, which = 1, caption = "Binomial Logit: Temp + Temp^2" )
plot(O.failed.binomial_temp.glm, which = 1, caption = "Binomial Logit: Temp")

plot(O.failed.binary_temp_sq.glm, which = 1, caption = "Binary Logit: Temp + Temp^2")
plot(O.failed.binary_temp.glm, which = 1, caption = "Binary Logit: Temp")
```


Question 6

6a) Model Evaluation

Answer 6a)

- Comparing our four major models, the binomial logistic regression model with temperature as the only explanatory variable have the lowest AIC, its residuals behaved best and its outliers are the least extreme. The model has given our data a fair representation. However, from the previous plots of predicted probabilities, the model estimates confidence bands that are too wide for us to assess risk of O-ring failure at 31F with any reasonable precision. Its coefficient estimates for temperature is also statistically insignificant. In addition to problems of dependence between observations, its usefulness in our study should be questioned.

- On the other hand, the two binary logistic models estimated statistical significant coefficients for temperature with similar (therefore more robust) values. Compared to the binomial models. They also estimate narrower confidence bands for predicted probabities of O-ring failure when temperature is 31F. Of the two binary models, we suggest to keep the more parsimonious model with "Temp" as the only

explanatory variable. Coefficient estimate for "Pressure" is not significant and it would help our main study only if the challenger had more launches before its accident. Although residuals in this model are more dispersed than its binomial counterpart, its estimates are more precise and useful for our study.

```
# regression table
suppressWarnings(stargazer::stargazer(O.failed.binomial.glm, O.failed.binomial_temp.glm, O.failed.binary
                    type = "text", title = "Comparison of Logistic Models",
                    star.cutoffs = c(0.05, 0.01, 0.001),
                    column.labels = c("binomial", "binomial", "binary", "binary"),
                    model.names = F
                    ))
```

```
##
## Comparison of Logistic Models
## ======================================================
## 		   Dependent variable:
## 		   -----------------------------------
## 		   O.failed.binomial  O.failed.binary
## 		   binomial  binomial binary   binary
## 		     (1)       (2)      (3)      (4)
## -------------------------------------------------------
## Temp              -0.098    -0.116  -0.229* -0.232*
## 		    (0.110)   (0.115)  (0.110) (0.108)
##
## Pressure           0.008            0.010
## 		    (0.019)           (0.009)
##
## Constant           2.520     5.085   13.292  15.043*
## 		    (8.541)   (7.477)  (7.664) (7.379)
##
## -------------------------------------------------------
## Observations        23        23      23      23
## Log Likelihood    -1.614    -1.601  -9.391  -10.158
## Akaike Inf. Crit.  9.229     7.203   24.782  24.315
## ======================================================
## Note:               *p<0.05; **p<0.01; ***p<0.001
```

6b) Interpret the main result of your model in terms of both odds and probability of failure.

Answer 6b)

- Our final model takes the form

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

where π refers to the probability of one or more O-ring failures of a Challenger rocket launch.

- Odds interpretation

$$OR = \frac{Odds_{T+c}}{Odds_T} = \frac{e^{\beta_0 + \beta_T(T+c)}}{e^{\beta_0 + \beta_T T}} = e^{c\beta_T} = e^{1*(-0.2322)} = 0.793$$

- The estimated odds of one or more O-ring failures per launch changes by 0.793 times, equivalent to decrease by 20.7%, with one unit of increase in temperature.

$$OR = \frac{Odds_{T+c}}{Odds_T} = \frac{e^{\beta_0 + \beta_T(T+c)}}{e^{\beta_0 + \beta_T T}} = e^{c\beta_T} = e^{(-7.057)*(-0.2322)} = 5.147$$

- The estimated odds of one or more O-ring failures per launch changes by 5.147 times, equivalent to increase by 414.8%, with one standard deviation of decrease (7.057F) in temperature.

$$OR = \frac{Odds_{T+c}}{Odds_T} = \frac{e^{\beta_0 + \beta_T(T+c)}}{e^{\beta_0 + \beta_T T}} = e^{c\beta_T} = e^{7.057*(-0.2322)} = 0.194$$

- The estimated odds of one or more O-ring failures per launch changes by 0.194 times, equivalent to decrease by 80.6%, with one standard deviation of increase (7.057F) in temperature.

- Probability of failure interpretation

  - Relative risk, the ratio between two probabilities, will always depend on holding specific value of temperature and specific value of increment. Since

$$\pi_i = \frac{e^{\beta_0 + \beta_T T_i}}{1 + e^{\beta_0 + \beta_T T_i}}$$

$$\pi_1 = \frac{e^{\beta_0 + \beta_T t}}{1 + e^{\beta_0 + \beta_T t}} \qquad \pi_2 = \frac{e^{\beta_0 + \beta_T(t+c)}}{1 + e^{\beta_0 + \beta_T(t+c)}}$$

$$RR = \frac{\pi_2}{\pi_1} = e^{c\beta_T} \cdot \frac{1 + e^{\beta_0 + \beta_T t}}{1 + e^{\beta_0 + \beta_T(t+c)}}$$

```
RR = function(mod.fit,t1,t2){
    c = t2 - t1
    beta.int = as.numeric(mod.fit$coefficients[1])
    beta.T = as.numeric(mod.fit$coefficients[2])
    exp(c*beta.T) * (1 + exp(beta.int + beta.T*t1)) / (1 + exp(beta.int + beta.T*(t1+c)))}

# increasing temperature from 31F to 51F
RR(O.failed.binary_temp.glm, t1 = 31, t2 = 53)
```

```
## [1] 0.9396154
```

```
# increasing temperature from 51F to 71F
RR(O.failed.binary_temp.glm, t1 = 31, t2 = 70)
```

```
## [1] 0.2300583
```

```
# increasing temperature from 71F to 91F
RR(O.failed.binary_temp.glm, t1 = 31, t2 = 81)
```

```
## [1] 0.02271217
```

- The probability of one or more O-ring failures per launch is 0.93 times as large, that is a decrease by 7%, if launching temperature is increased from 31F, the temperature at which the Challenger had its accident, to 53F the lowest temperature among its previous launches.

- The probability of one or more O-ring failures per launch is 0.23 times as large, that is a decrease by 77%, if launching temperature is increased from 31F to 70F, the median temperature among its previous launches.

- The probability of one or more O-ring failures per launch is 0.02 times as large, that is a decrease by 98%, if launching temperature is increased from 51F to 71F, the maximum temperature among its previous launches.

6c) Plot the main effect of your final model with the y-axis being probability of failure and x-axis being temperature.

Answer 6c)

- Predicted probability of O-ring failure for our final model peaks once temperature drops below 40F. Negative slope of this probability curve is the steepest (predicted probability drops most drastically) at predicted probability of 0.5 and temperature between 60F and 70F. The probability curve then smooths off until it hits our maximum temperature of 81F in dataset, when predicted probability almost drops to zero.

- The changing width of confidence band has already been covered in section 5c).

```r
# base plot of observations
y = df$O.ring/6

plot(x = df$Temp, y = jitter(y,0.02), xlab = "Temperature in F",
     ylab = "Probability of O-ring failure",
     panel.first = grid(col = "gray", lty = "dotted"), ylim = c(0,1), xlim = c(30,81),
     main = "Predicted Probability of O-ring failure vs Temperature ")

curve(expr = predict(object = O.failed.binary_temp.glm,
      newdata = data.frame(Temp= x), type = "response"),
      col = "blue", xlim = c(31,81), add = T)

# Plot CI band
curve(expr = ci.pi(newdata = data.frame(Temp = x), mod.fit.obj = O.failed.binary_temp.glm, alpha = 0.05)
      xlim = c(31,81), add = T, lty = "dotted", col = "blue")
curve(expr = ci.pi(newdata = data.frame(Temp = x), mod.fit.obj = O.failed.binary_temp.glm, alpha = 0.05)
      xlim = c(31,81), add = T, lty = "dotted", col = "blue")

# Legend
legend(x = 28, y = 0.4, legend = c("Binary Logit model", "95% Confidence Interval"),
     lty = c("solid","dotted"), col = c("blue"), bty = "n")
legend(x = 31, y = 0.5, legend = c("observations"),
     pch = 1, bty = "n")
```
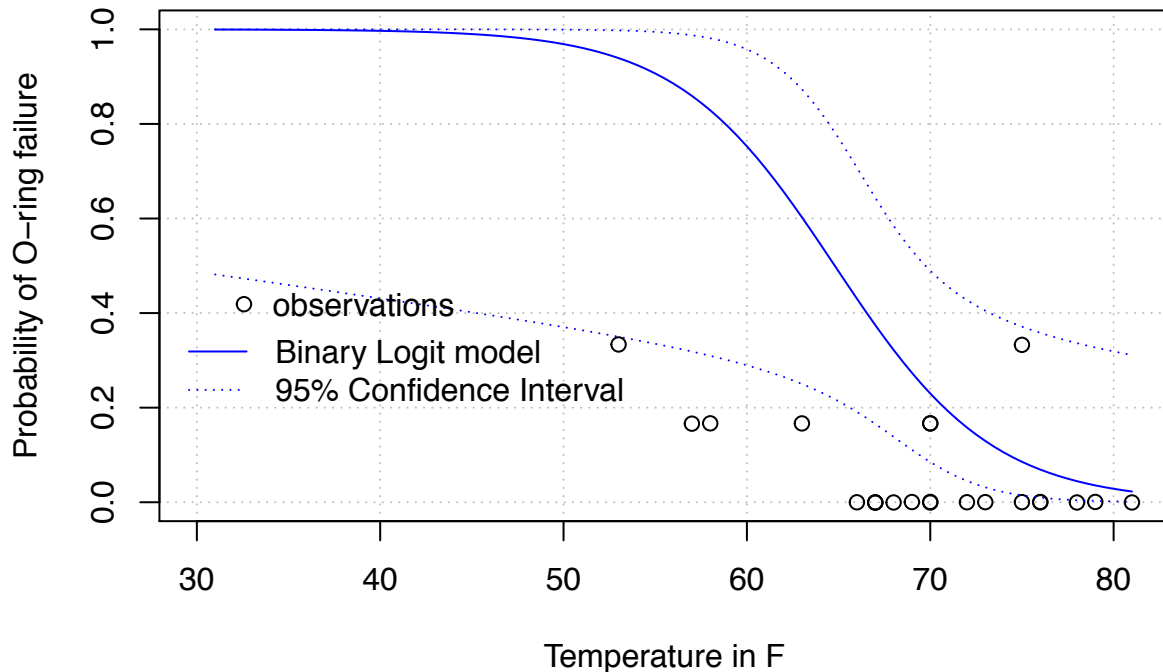
Predicted Probability of O–ring failure vs Temperature

Conclusion

The failure of an O-ring on the space shuttle Challenger's booster rockets led to its destruction in 1986. In this lab assignment, we built four major models using data on previous space shuttle launches, Dalal et al. (1989) to reexamine the probability of an O-ring failure as a function of temperature at launch and combustion pressure: bionomial logistic regression models and binary logistic regression models. Temperature is statistically significant while Pressure is not as important from the analysis. While binary logistic regression models show more robust values for predictors with narrower confidence bands for predicted probabilities of O-ring failure, both binomial and logistic regression models lead to a similiar conclusion that failure of an O-ring(s) is certain at temperature before 40 F and mostly all 6 O-rings would fail during flight.

Although hindsight will always be 20/20, Challenger dataset leave a lot to be desired. There are concerns with variables in the dataset due to bias, interactions, or with limited value while other "not available" variables could play a role in predicitng O-ring failure(s).