Analysis of Challenger Shuttle Failure

Stone Jiang

This report will use logistic regression to analyze partial data available before the Challenger Space Shuttle failure. We will focus on both the predictive power of logistic regression, and inferential procedures around coefficient interpretation and confidence interval generation. The central question are:

1. Should the launch have been delayed for a later date with higher temperature?
2. Given past data, what does the LR predict as the odds and probability, as associated confidence intervals, of shuttle failure?

Notably, we will implement a parametric bootstrapping procedure for estimating the confidence interval of logistic regression parameters, and compare its interval width to both the the Wald and profile LR confidence intervals.

# Investigation of the 1989 Space Shuttle Challenger Accident

```r
library(knitr)
library(Hmisc)
library(car)
library(dplyr)
library(ggplot2)
library(mcprofile)
library(patchwork)
#Set up global knitr option so that text does not go off of the page
opts_chunk$set(tidy.opts=list(width.cutoff=60),tidy=TRUE)
```

Conduct a thorough EDA of the data set.

```r
challenger <- read.csv("data/challenger.csv")
describe(challenger)
```

```
## challenger 
## 
##  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
## ---------------------------------------------------------------------------
```

We've provided a citation at the end of the document. All references (such as when we refer to the "authors") in this report are with respect to that citation.

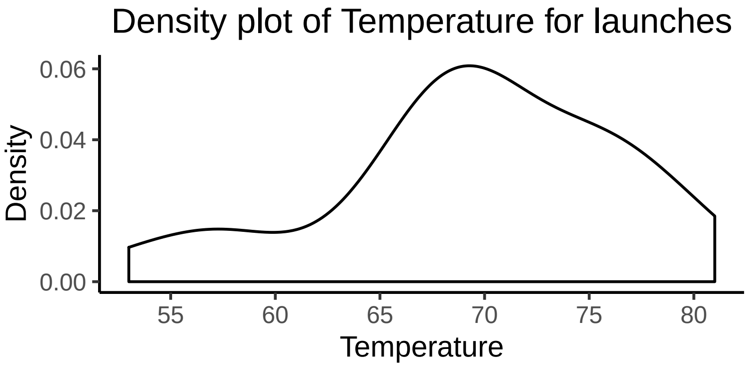
The tabular form of the data is shown above. First of all, we see that there are no missing values and all variables are numeric. "Flight" is just a label and contains no predictive value, so we will not use this in our models. "Number," representing the number of O-rings for each flight, is 6 across the board, which is correct since we know that each flight has 6 O-rings.

The dependent variable is "O.ring"" variable, which ranges from 0 to 2, represents the number of O.rings that failed for each launch. These values are reasonable since there are at most 6 O-rings per launch. We can tabulate the data and see that most launches resulted in 0 failures, while 5 launches resulted in a single failure. There were 2 launches which had 2 failures. Clearly, there are more negative cases (no failure) than positive cases, and there are more single O-ring failures than double failures per launch.

We next examine the Temperature variable in more detail:

The current range is from 53 degrees to 81 degrees, which is significantly higher than 31 degrees, the temperature on the day of the launch. While most temperatures had a single value, 70 degrees was by far the most common, representing the temperature for 4 distinct test launches, followed by 67 degrees, with 3 distinct test launches. We can see the distribution a little better on a density plot:

```
ggplot(challenger, aes(x = Temp)) + geom_density() + labs(x = "Temperature",
    y = "Density", title = "Density plot of Temperature for launches") +
    theme_classic() + theme(plot.title = element_text(hjust = 0.5)) +
    scale_x_continuous(breaks = seq(50, 85, 5))
```



We can see that the majority of the launches happened above 65 degrees, making this distribution left skewed, with a small tail for lower temperatures. We anticipate that predictions at lower temperatures may have larger confidence intervals as a result of fewer data points.

Next, we examine the Pressure variable.

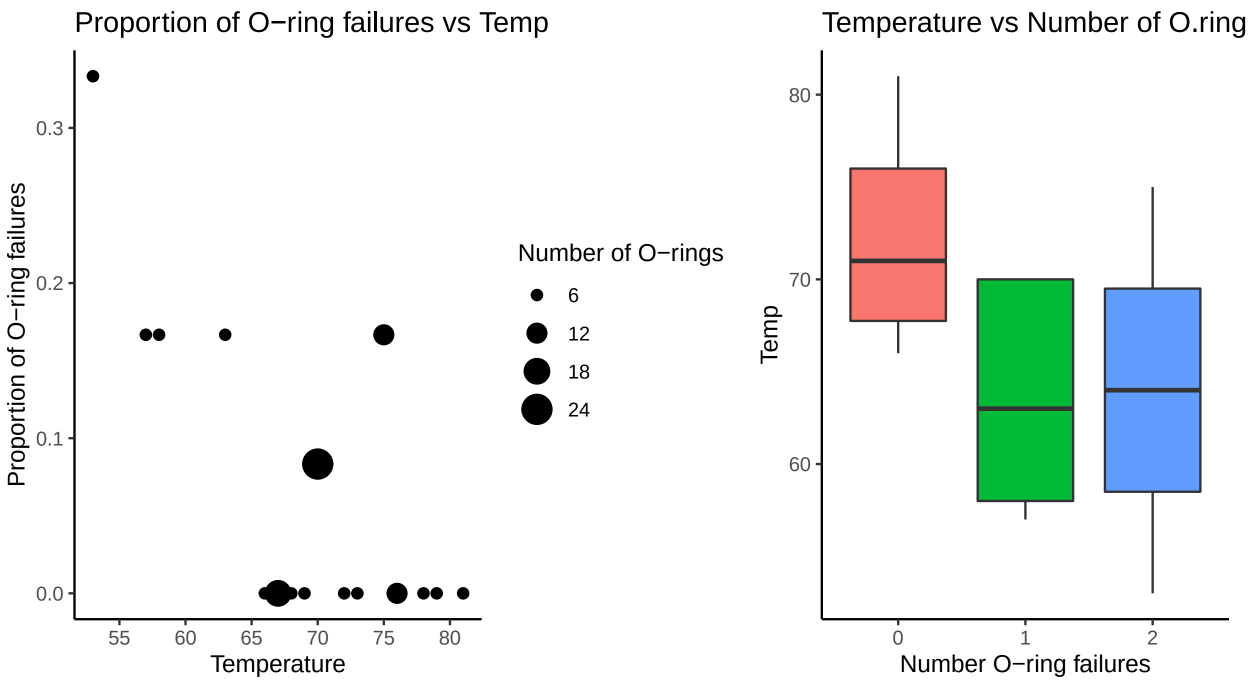
Pressure has three values, which is by design. The values correspond to the leak pressured applied to the O-ring. Values of 50psi correspond to earlier tests, before the discovery that a putty used to construct the motor could withold this pressure, making this test invalid for testing O-rings (since the O-rings would not feel the pressure). This was briefly raised to 100psi for 2 tests, and for the majority of tests which happened in 1984 and later, was conducted at 200psi. The increase in the Pressure test could have caused "blow holes" in the putty, which could contribute to O-ring erosion and failure by letting hot air through, which is the reason that Pressure of the test could have an impact on O-ring failure. Note that we will treat Pressure variable as a numeric variable even though there are only 3 categories because we want our model to be applicable to more than just 3 pressures. For example, we may decide later to conduct the test at 300psi. If we treat Pressure as categorical, the model would no longer to applicable.

Since we will be looking at O.ring failures with respect to temperature and pressure, we will visualize these relations both individually and together. We first look at a scatter plot of the proportion of failures for each temperature, rather for each launch. The size of our scatter points will be proportional to the number of trials at the temperature. To do this, we group the data by temperature (irrespective of Flight), and calculate both the number of failures and number of total O-rings (trials) at each temperature.

```
# Generate dataframe aggregated by temperature
temp.agg <- challenger %>% group_by(Temp) %>% summarise(O.ring = sum(O.ring),
    Number = sum(Number), prop = O.ring/Number)
rbind(temp_prop_cor = with(temp.agg, cor(prop, Temp)))
```

```
##                     [,1]
## temp_prop_cor -0.7305724
```

```
# Generate a plot for proportion of failures versus
# temperature
pa = ggplot(data = temp.agg, aes(x = Temp, y = O.ring/Number)) +
    geom_point(aes(size = factor(Number))) + labs(x = "Temperature",
    title = "Proportion of O-ring failures vs Temp", y = "Proportion of O-ring failures",
    size = "Number of O-rings") + theme(plot.title = element_text(hjust = 0.5)) +
    theme_classic() + scale_x_continuous(breaks = seq(50, 85,
    by = 5))
pb = ggplot(challenger, aes(factor(O.ring), Temp)) + geom_boxplot(aes(fill = factor(O.ring)),
    show.legend = FALSE) + ggtitle("Temperature vs Number of O.ring failure") +
    theme(plot.title = element_text(lineheight = 1, face = "bold")) +
    labs(x = "Number O-ring failures", y = "Temp") + theme_classic()
pa + pb
```



First of all, the correlation between Temperature and Proportion of O-ring failures is high, at -0.73. The negative sign suggests lower temperature means higher proportion of O-ring failures. It also appears that the cases that had no failures were at >65 degrees. The boxplot similarly suggests that for launches with 0 failures, the distribution of temperature appears to be higher than for launches with more than 0 failures. Furthermore, the left plot suggests that as the temperature decreases, the proportion of incidents tends to increase (higher proportion of failure was the launch at 53 degrees), even though there is a smaller number of O-rings tested at these lower temperatures. The exceptions to this is at 70 and 75 degrees. At 70 degrees, there were 24 trials. Let's say that the trend holds and the true probability of failure is lower at 70 degrees than at 66 degrees, which had 6 trials. We shouldn't be surprised to observe that the absolute number of failure cases is >0 at 70 degrees even when we do not observe any at 63 degrees. For the 75 degrees data point, we extract them below:

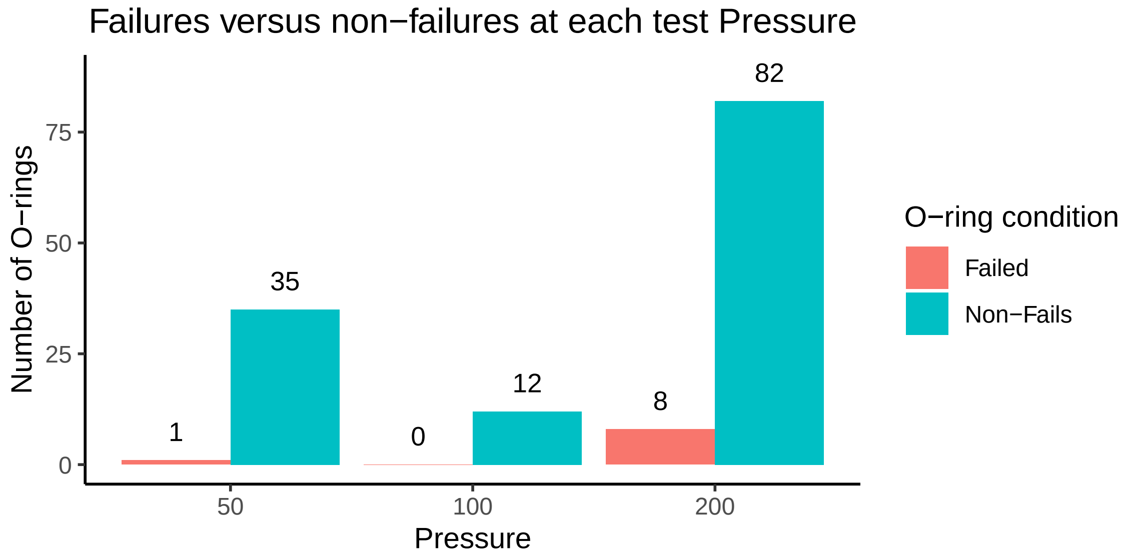
```
challenger %>% filter(Temp == 75)
```

```
##   Flight Temp Pressure O.ring Number
## 1     16   75      200      0      6
## 2     21   75      200      2      6
```

We see that all of the failures for this temperature came from a single launch 21. This launch appears to be an outlier in our dataset. Without further background on the circumstances of the launch, we cannot make conclusions about whether this is a representative data point. For example, omitted variables such as operational mistakes, could have caused this shuttle to failure at 2 O-ring locations.

Next, we look at the relationship between the Pressure and O-ring failure variables. Since Pressure only takes on 3 values, we will show the number of O-rings that failed versus the number of O-rings that did not failure for each pressure using a barplot.

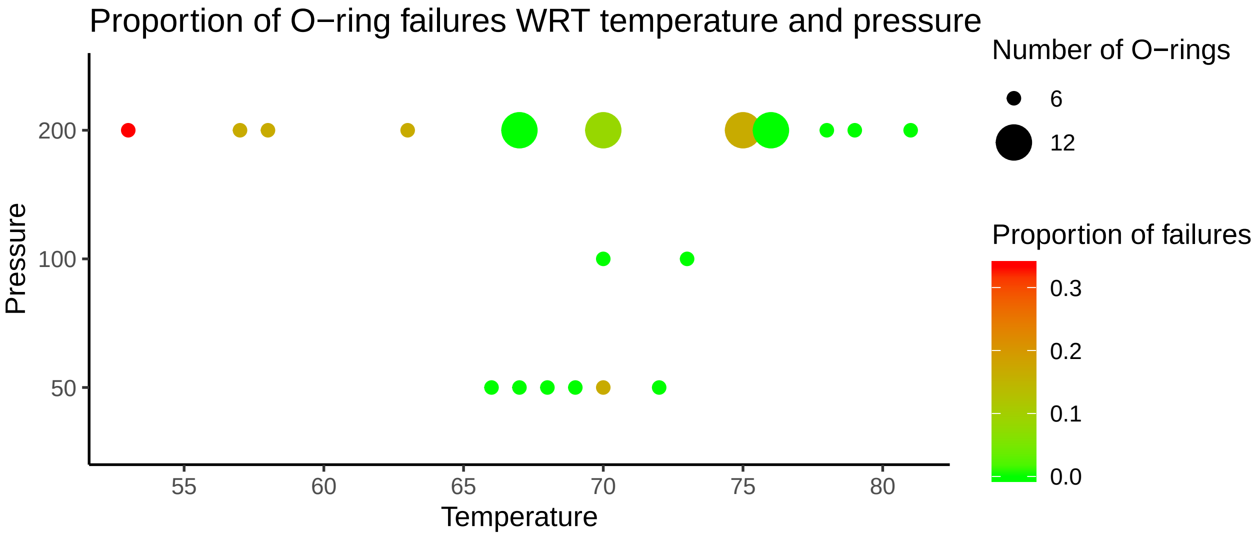
```r
# Generate number of failures for each Pressure
pressure.agg <- challenger %>% group_by(Pressure) %>% summarise(O.ring = sum(O.ring),
    Number = sum(Number), cond = "Failed")
# Generate number of non-failures for each Pressure, and join
# to dataframe
pressure.agg <- pressure.agg %>% mutate(cond = "Non-Fails", O.ring = Number -
    O.ring) %>% rbind(pressure.agg)
ggplot(data = pressure.agg, aes(x = factor(Pressure), y = O.ring,
    fill = cond)) + geom_bar(stat = "identity", position = position_dodge()) +
    geom_text(aes(label = O.ring), vjust = -1, color = "black",
        position = position_dodge(0.9), size = 3.5) + ylim(0,
    88) + labs(x = "Pressure", y = "Number of O-rings", fill = "O-ring condition",
    title = "Failures versus non-failures at each test Pressure") +
    theme_classic() + theme(plot.title = element_text(hjust = 0.5)) +
    scale_x_discrete(breaks = c(50, 100, 200))
```



We will not calculate a correlation between pressure and failures since there would only be 3 data points (one for each pressure). But we can see from this plot that there are a lot more failures occurring at 200psi than 50psi or 100 psi. There were also a lot more tests conducted at 200psi. Flight 21, which was a problematic outlier on the temperature versus O-ring analysis, is part of the 200psi set, which means the value 8 could be artifically inflated if that flight was not a representative launch. The general trend still is that 200psi appears to have more failures than lower psi values.

Finally, we will look at each O-ring as a single observation, and color by the proportion of failures on a 2D plot of Temperature and Pressure. We will also show the number of trials at each condition with the size of the dot. To do this, we group based on Temperature and Pressure, and aggregate the number of failures, as well as number of trials.

```
temp.pres.agg <- challenger %>% group_by(Temp, Pressure) %>%
    summarise(O.ring = sum(O.ring), Number = sum(Number))
ggplot(temp.pres.agg, aes(x = Temp, y = factor(Pressure))) +
    geom_point(aes(size = factor(Number), color = O.ring/Number)) +
    labs(title = "Proportion of O-ring failures WRT temperature and pressure",
        x = "Temperature", y = "Pressure", size = "Number of O-rings",
        color = "Proportion of failures") + theme(plot.title = element_text(hjust = 0.5)) +
    theme_classic() + scale_x_continuous(breaks = seq(50, 85,
    by = 5)) + scale_colour_gradientn(colours = c("green", "red"))
```



From this trivariate analysis, the clearest trend is stil that lower temperatures tend to have higher rates of failure. While Pressure on its own seemed to influence rates of failure, it appears that all tests at lower pressures were conducted at higher temperatures (>65). As a result, conditional on the fact that Temperature is an explanatory variable, the trend previously explained by Pressure is not entirely orthogonal to the variation in failure observed as a function of temperature. Overall this analysis seems to suggest that Temperature and Pressure might both be important, but that the importance of Pressure is somewhat diminished with Temperature in the same model.

The following is from Question 4 of Bilder and Loughin Section 2.4 Exercises (page 129):

(a) 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.

There are two ways to view why the independence assumption is necessary. The first is to treat each O-ring (regardless of which flight it is on) as a Bernoulli random variable. Here, the independence assumption is necessary for the MLE estimation. MLE assumes that each observation is independent and identically distributed so that the likelihood function (and joint distribution of the observations) is a product of the marginal distributions. The second way is to treat each launch as a separate Binomial random variable. The Binomial model requires that each of the 6 O-rings are iid, and MLE estimation using the Binomial requires that the launches are also iid. Both views lead to the same model that all O-rings, including those on the same launch, are independent.

This is potentially problematic because O-rings, conditioned on the same launch, could have dependencies due to being on the same shuttle. For example, the failure of one ring could put more strain on the others on the same shuttle, resulting in a higher chance that subsequent failures also occur. Additionally, putty is used to protect the O-rings from heat, so an abnormality in the manufacturing process (such as uneven distribution of putty) could make some O-rings of a particular shuttle more vulnerable than others due to their positioning on the shuttle. The authors noted later that using the binary model (where response is whether at least 1 O-ring failed rather than number or proportion of failures) no longer requires independence assumption of each joint.

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

To answer the questions, we will use the binomial model, in which we assume that all of the O-rings within the same launch is independent of each other. This allows each launch to have 6 trials (modeled as a binomial distirbution), and since each launch is independent of the next, we can form the likelihood as a product of binomials. The other model used in the paper was the binary model, where, instead of looking at per O.ring failure, the authors looked at the per launch failure. This was done by treating each launch as independent, and the response Y was whether there was any O.ring failures at all. In other words, Y = 0 if the number of failures was 0, and Y = 1 if the number of failures was greater than 0. The authors felt that this model was more robust to the actual number of incidents. The authors also showed that the two model agreed quite closely, that the lack of fit criterial G^2 was quite good for the binomial model, and that this likeness gave "more confidence in the binomial-logist model (pg 949)." In addition, given the independent assumption of the binomial model, the probability of failure (p^*) in the binary model (in other words, at least 1 O-ring failure) can be converted to the probability of failure in the binomial model (p) exactly as:

$$p^*(t) = 1 - (1 - p(t))^6$$

This is because the probability of at least 1 failure is 1 minus the probability of no failures. $1 - p(t)$ is the probability of no failure for one trial, and raising this to the 6th gives probability of no failures in 6 trials. Finally, 1 minus this quantity gives the probability of at least 1 failure. The major downside for the binary model is that some information is lost due to the fact that there are some launches in which the number of failures > 1, which though the authors felt the information loss was not too significant. For these reasons, and the fact that the authors used the binomial model up until the analysis of nozzle-joint data, we will fit the binomial model as well.

Letting $\pi_{failure}$ equal the probability that an O-ring fails, the model we want to fit is:

$$\log \frac{\pi_{failure}}{1-\pi_{failure}} = \beta_0 + \beta_1 * \text{Temp} + \beta_2 * \text{Pressure}$$

```r
# To produce the model with count data, we can use the
# weights column
mod.fit <- glm(O.ring/Number ~ Temp + Pressure, weights = Number,
    data = challenger, family = binomial(link = "logit"))
mod.fit
```

```
##
## Call:  glm(formula = O.ring/Number ~ Temp + Pressure, family = binomial(link = "logit"),
##     data = challenger, weights = Number)
##
## Coefficients:
## (Intercept)         Temp     Pressure
##    2.520195    -0.098297     0.008484
##
## Degrees of Freedom: 22 Total (i.e. Null);  20 Residual
## Null Deviance:       24.23
## Residual Deviance: 16.55     AIC: 36.11
```

The numeric form of our model is:

$$\log \frac{\pi_{failure}}{1-\pi_{failure}} = 2.520 - 0.0983 * \text{Temp} + 0.0085 * \text{Pressure}$$

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

To do this, we perform the LRT test using Anova at a significance level of 0.05. The null and alternative hypothesis for each coefficient is as follows.

For Temperature, given that Pressure is in the model:

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

For Pressure, given that Temperature is in the model:

$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$

```r
Anova(mod.fit)
```

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

Based on the profile LR test, Pressure is not significant given that Temperature is in the model since p>0.05. We therefore fail to reject our null hypothesis for Pressure. However, Temp is statistically significant given that Pressure is in the model p>0.05, so we reject H_0 for Temperature and claim that its effect on O-ring failure is significant.

(d) 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?

There are only 2 variables, and Pressure has a large p-value > 0.05, meaning it is not statistically significant given Temp was in the model, so the authors chose to remove it. One problem might be that all we know is that Temperature is significant given that Pressure is in the model, so removing Pressure in theory can eliminate the significance of Temperature. Since we only have two variables here, this is probably unlikely. The bigger issue is that alternative forms of Pressure, such as quadratic terms, interactions with Temperature, transformations, etc, might indeed be significant. To address this we could for example use glmulti to exhaustively search for all models with higher order terms and interactions in order to increase the complexity of the model. Since we won't do this here, we will use Temperature only in our final model. In addition, Pressure might not be statistically significant but that does not imply that it has no explanatory power. It may still be more a predictive model to include both effects (Pressure and Temperature).

The following are from Question 5 of Bilder and Loughin Section 2.4 Exercises (page 129-130):

Continuing Exercise 4, consider the simplified model $logit(\pi) = \beta_0 + \beta_1 Temp$, where π is the probability of an O-ring failure. Complete the following:

(a) Estimate the model.

The form of the model is provided in the problem statement. We wish to estimate using only the temperature variable.

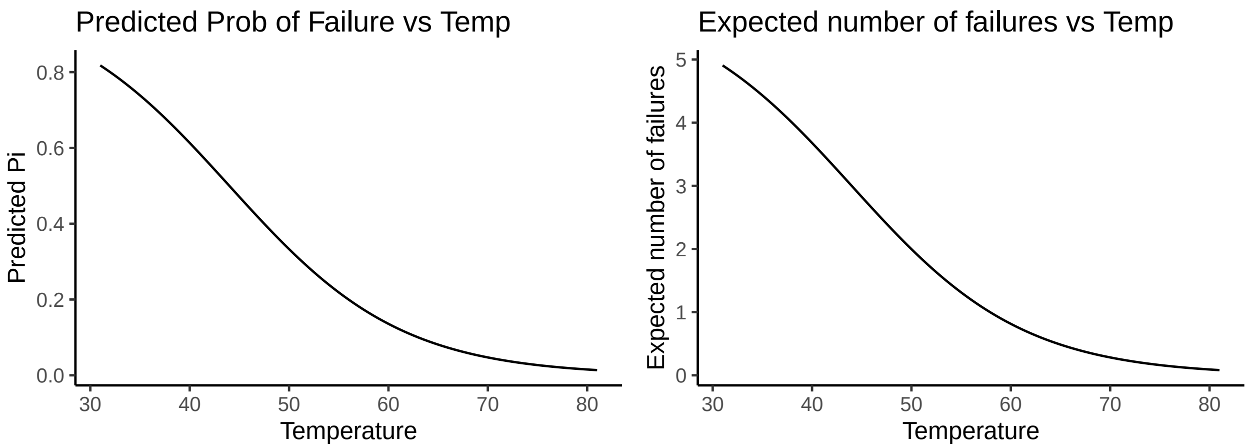
```
mod.fit <- glm(O.ring/Number ~ Temp, weights = Number, data = challenger,
    family = binomial(link = "logit"))
mod.fit
```

```
##
## Call:  glm(formula = O.ring/Number ~ Temp, family = binomial(link = "logit"),
##     data = challenger, weights = Number)
##
## Coefficients:
## (Intercept)         Temp
##      5.0850      -0.1156
##
## Degrees of Freedom: 22 Total (i.e. Null);  21 Residual
## Null Deviance:       24.23
## Residual Deviance: 18.09     AIC: 35.65
```

$$\log \frac{\pi_{failure}}{1 - \pi_{failure}} = 5.085 - 0.1156 * \text{Temp}$$

(b) Construct two plots: (1) π vs. Temp and (2) Expected number of failures vs. Temp. Use a temperature range of 31° to 81° on the x-axis even though the minimum temperature in the data set was 53°.

```
# Construct the temperature range
Temp <- data.frame(Temp = seq(31, 81, by = 0.5))
# The pi is the predicted
predicted_pi <- predict(mod.fit, Temp, type = "response")
# The expected number is 6 * pi, since we're assuming each
# O-ring is independent for each launch
expected.num.failure <- predicted_pi * 6
# ggplot expects a dataframe
data.plotting <- data.frame(Temp = Temp$Temp, predicted_pi = predicted_pi,
    expected.num.failure = expected.num.failure)
p1 <- ggplot(data = data.plotting, aes(x = Temp, y = predicted_pi)) +
    geom_line() + labs(title = "Predicted Prob of Failure vs Temp",
    x = "Temperature", y = "Predicted Pi") + theme_classic()
p2 <- ggplot(data = data.plotting, aes(x = Temp, y = expected.num.failure)) +
    geom_line() + labs(title = "Expected number of failures vs Temp",
    x = "Temperature", y = "Expected number of failures") + theme_classic()
p1 + p2
```



(c) Include the 95% Wald confidence interval bands for π on the plot. Why are the bands much wider for lower temperatures than for higher temperatures?

To construct the Wald confidence interval for π, we note that the Normal approximation is better for $\beta_0 + \beta_1 * Temp$ than it is for the expression for π itself, so we will first build the CI for the linear predictor, then transform that into the CI for π. The Wald CI for the linear predictor is:

$$\hat{\beta}_0 + \hat{\beta}_1 * Temp \pm Z_{1-\frac{\alpha}{2}} * \sqrt{V\hat{a}r(\hat{\beta}_0 + \hat{\beta}_1 * Temp)}$$

The variance term is:

$$V\hat{a}r(\hat{\beta}_0 + \hat{\beta}_1 * Temp) = V\hat{a}r(\hat{\beta}_0) + Temp^2 V\hat{a}r(\hat{\beta}_1) + 2 * Temp * C\hat{o}v(\hat{\beta}_0, \hat{\beta}_1)$$

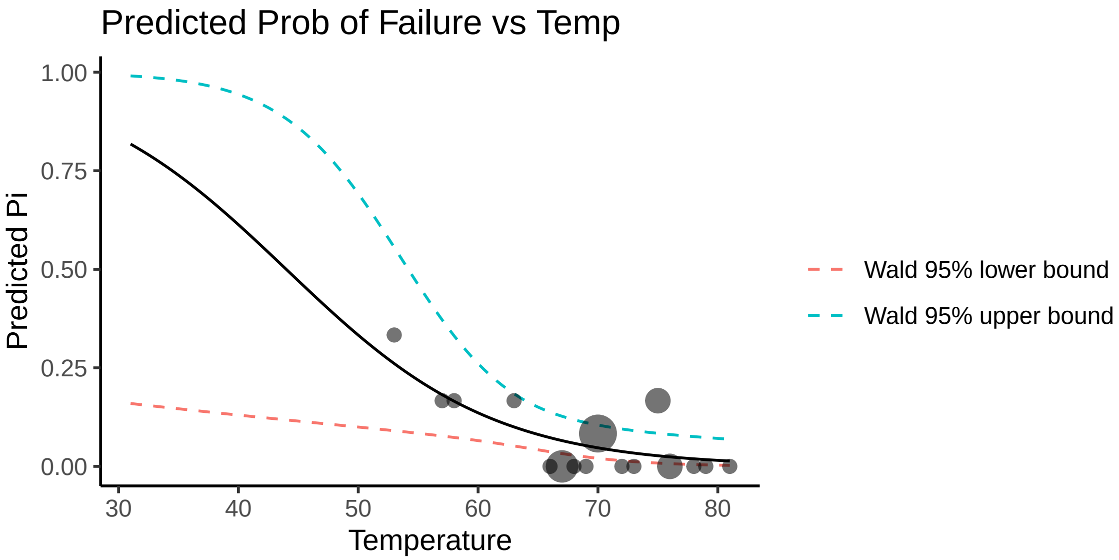
The variance term can be derived from the predict function.

Note that we will plot with the observed proportion of failures for each temperature (size of dot scaled by the number of observations) overlaid on top.

```r
prediction <- predict(mod.fit, Temp, type = "link", se = TRUE)
var <- prediction$se.fit
lp <- prediction$fit
alpha = 0.05

CI.lp.lower <- lp + qnorm(alpha/2) * var
CI.lp.upper <- lp + qnorm(1 - alpha/2) * var
CI.lp <- cbind(lower = CI.lp.lower, upper = CI.lp.upper)
CI.pi <- exp(CI.lp)/(1 + exp(CI.lp))
CI.plotting <- data.frame(CI.pi, predicted_pi = predicted_pi,
    Temp = Temp)

ggplot(data = CI.plotting, aes(x = Temp)) + geom_line(aes(y = predicted_pi)) +
    geom_line(aes(y = lower, color = "Wald 95% lower bound"),
        linetype = "dashed") + geom_line(aes(y = upper, color = "Wald 95% upper bound"),
    linetype = "dashed") + geom_point(data = temp.agg, aes(y = prop,
    size = factor(Number), alpha = 0.01), show.legend = FALSE) +
    labs(title = "Predicted Prob of Failure vs Temp", x = "Temperature",
     y = "Predicted Pi", color = "", alpha = "Actual observations") +
  theme_classic()
```



The bands are much wider for lower temperatures because there were much fewer observations there. Most of our observations are clustered around 65-80 degrees. In fact, as can be seen on the graph, we have no observations below 53 degrees, so it would make sense that the model is much more uncertain (wider CI means more uncertainty in the estimate) for lower temperatures values. Mathematically, small number of observations results in large standard errors of the predictions at those values, leading to larger confidence intervals.

(d) The temperature was 31° 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.

We will first look at the Wald CI already constructed, then construct the profile LR CI, and if the results are similar and the profile LR CI gave no warnings, then we will use the LR CI. We do both at 95% confidence level.

```
# The wald was previously calculated
wald.info <- CI.plotting %>% filter(Temp == 31) %>% mutate(Estimate = predicted_pi)
wald.ci = wald.info[0:3]
names(wald.ci) = c("lower", "upper", "Estimate")

# To calculate the profile LR CI, we will use the mcprofile.
K <- matrix(c(1, 31), nrow = 1)
linear.combo <- mcprofile(mod.fit, CM = K)
temp.LR.ci <- confint(linear.combo, level = 0.95)
lr.estimate <- exp(temp.LR.ci$estimate)/(1 + exp(temp.LR.ci$estimate))
lr.ci <- exp(temp.LR.ci$confint)/(1 + exp(temp.LR.ci$confint))

cis <- rbind(wald = wald.ci, profile = cbind(lr.ci, lr.estimate))
cis
```

```
##               lower     upper  Estimate
## wald    0.1596025 0.9906582 0.8177744
## profile 0.1418508 0.9905217 0.8177744
```

The estimated probability of failure is 0.8178. We can see that the CIs are very similar to each other. The wald CI lower bound is a bit higher than the profile LR, but we also know that the Wald CI typically undershoots the true significant level, so it would make sense that the true CI should be a bit wider. As a result, we will use the profile LR for interpretation. With 95% confidence, the estimated probability of each O-ring failure is between 0.142 and 0.991 at 31 degrees.

The primary assumption of the Wald CI is discussed above, that the linear predictor is normally distributed. The profile likelihood requires regularity conditions to ensure that the ML estimators are unique. In the context of this problem, we are assuming that the model holds for even ranges of temperature for which there are no observations. The temperature is of 31 is far outside the range of observations, so we are assuming that the estimated relationship can be extrapolated to these low temperatures. With a single observation at say 30 degrees, our estimates can greatly change as that point will have high leverage and potentially high influence.

(e) Rather than using Wald or profile LR intervals for the probability of failure, Dalal et al. (1989) use a parametric bootstrap to compute intervals. Their process was to (1) simulate a large number of data sets (n = 23 for each) from the estimated model of Temp; (2) estimate new models for each data set, say and (3) compute at a specific temperature of interest. The authors used the 0.05 and 0.95 observed quantiles from the simulated distribution as their 90% confidence interval limits. Using the parametric bootstrap, compute 90% confidence intervals separately at temperatures of 31° and 72°.

```
# Simulate a large number of datasets (n = 23 for each) from
# the estimated model of temp We need a reasonable range of
# temperatures from which to simulate the data To sample
# temperature, we will sample with replacement from the
# initial dataset We will use parametric bootstrapping by
# default as stated in the problem

single_bootstrap <- function(mod, n, size, temp, btype = "parametric") {
    # sample temperatures from the dataset
    Temp <- data.frame(Temp = sample(challenger$Temp, n, replace = TRUE))
    # generated the predicted pi.hat value for each sampled
    # temperature with the beta parameters of our original model
    pi <- predict(mod, newdata = Temp, type = "response")
    if (btype == "parametric") {
        # using the generated pi.hat values, run binomial sampling to
        # generate number of failures for each of the pi values. This
        # is the parametric bootstrap procedure where we assume that
        # each example is generated from the binomial
        samples <- rbinom(n = n, size = size, prob = pi)
        dataset <- data.frame(Temp, O.ring = samples, Number = size)
    } else if (btype == "nonparameteric") {
        dataset <- challenger[sample(nrow(challenger), 23, replace = TRUE),
            ]
    } else {
        stop("Choose parameteric and nonparameteric for type of bootstrapping")
    }
    # fit the model with the dataset
    mod.bootstrap <- glm(O.ring/Number ~ Temp, weights = Number,
        data = dataset, family = binomial(link = "logit"))
    # get the estimated pi value at the specific temperature of
    # interest
    pi.hat <- predict(mod.bootstrap, newdata = data.frame(Temp = temp),
        type = "response") %>% unname
    return(pi.hat)
}
samples_per_temperature <- 500
temps <- c(31, 72)
simulated_data <- matrix(, nrow = 2, ncol = samples_per_temperature)

for (temp_ind in c(1, 2)) {
    for (i in seq(1:samples_per_temperature)) {
        pi.hat <- single_bootstrap(mod.fit, 23, 6, temps[temp_ind],
            btype = "parametric")
        simulated_data[temp_ind, i] <- pi.hat
    }
}
quantile_0.05 <- rbind(temp31 = quantile(simulated_data[1, ],
    0.05), temp72 = quantile(simulated_data[2, ], 0.05))
quantile_0.95 <- rbind(temp31 = quantile(simulated_data[1, ],
    0.95), temp72 = quantile(simulated_data[2, ], 0.95))
quantiles <- data.frame(quantile_0.05, quantile_0.95)
names(quantiles) = c("90%CI_lower", "90%CI_upper")
quantiles
```

```
##        90%CI_lower 90%CI_upper
## temp31  0.08742237  0.99351004
## temp72  0.01063221  0.07173748
```

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

We will fit a model with both a quadratic term and order 1 term for temperature and evaluate by the LRT to see whether this model is significantly different than a model with only the order 1 term. We test at a significance level of 0.05. Our hypothesis is stated below:

$$H_0 : logit(\pi) = \beta_0 + \beta_1 * \text{Temp}$$
$$H_1 : logit(\pi) = \beta_0 + \beta_1 * \text{Temp} + \beta_2 * \text{Temp}^2$$

```
mod.fit.sqrd <- glm(O.ring/Number ~ Temp + I(Temp^2), weights = Number,
    data = challenger, family = binomial(link = "logit"))
anova(mod.fit, mod.fit.sqrd, test = "LRT")
```

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

```
## Model 1: O.ring/Number ~ Temp
## Model 2: O.ring/Number ~ Temp + I(Temp^2)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        21     18.086
## 2        20     17.592  1   0.4947   0.4818
```

Based on the profile LR test, we fail to reject H0 and conclude that the Temp^2 is not statistically significant given that the order 1 term Temp is in the model (no sign of a quadratic relationship), since the p-value >>0.05. We can also see which model is better based on the AIC.

```r
aic.h0 <- AIC(mod.fit)
aic.ha <- AIC(mod.fit.sqrd)
aic <- cbind(h0 = aic.h0, ha = aic.ha)
row.names(aic) = c("aic")
aic
```

```
##           h0       ha
## aic 35.64654 37.15184
```

Since Temp^2 is not significant given that Temp is in the model, and the AIC is lower for the model in H0, we do not want to include the quadratic term in our final model.

With the same set of explanatory variables in your final model, estimate a linear regression model. Explain the model results; conduct model diagnostic; and assess the validity of the model assumptions. Would you use the linear regression model or binary logistic regression in this case? Explain why.

We will only then fit the model with Temperature since Pressure is not significant, and Temperature squared is not needed in the logistic regression model. Note we make the same assumptions as in the logistic regression model, which is that each O.ring, regardless of the launch, is independent of each other. As a result, we need to separate out each example. Each launch will have 6 (assumed to be) independent examples, with n of them marked as 1, and 6-n marked as 0, where n is the number of failures for that particular launch. Each O-ring is then treated as a separate example.

Another way to approach this problem is to treat each launch as independent, and calculate the fraction of failures per launch. In this case, each launch is a data point. Both are legitimate ways to perform modeling; however, this second method is not consistent with the assumptions made in the binomial model, which is that every O-ring is independent. As a result, we will stick to the former.

Let π be the probability of O.ring failure as a function of Temperature. The model we want is:

$$\pi = \beta_0 + \beta_1 * \text{Temp}$$

```
# Expand each launch into 6 separate data points For each
# launch, number of points with O.ring outcome = 1 is number
# of failures for each launch
challenger.expanded <- data.frame(matrix(nrow = 0, ncol = length(challenger)))
colnames(challenger.expanded) <- colnames(challenger)

for (rname in row.names(challenger)) {
    # Extract the row information
    r <- challenger[rname, ]
    for (times in seq(r[["Number"]])) {
        challenger.expanded <- rbind(challenger.expanded, r)
        # This decrements O.ring by 1, and keeps the minimum at 0
        r["O.ring"] = as.numeric((r["O.ring"] - 1) > 0)
    }
}
# First, we make all O-ring values that were 2 the number 1
# instead
challenger.expanded$O.ring <- as.numeric(challenger.expanded$O.ring >
    0)
# Fit the model of fraction of failures to each temperature
mod.lm <- lm(O.ring ~ Temp, data = challenger.expanded)
mod.lm
```

```
##
## Call:
## lm(formula = O.ring ~ Temp, data = challenger.expanded)
##
## Coefficients:
## (Intercept)         Temp
##     0.616402    -0.007923
```

The model we have fit is:

$$\pi = 0.6164 - 0.0079 * \text{Temp}$$

This model interpretation is that for a 1 unit increase in Temperature, the probability of O.ring failure decreases by 0.0079, regardless of where the initial temperature starts at. At 0 degrees, the probability of failure is 0.6164.

Now we assess model validity and fit by examining the CLM assumptions.

CLM 1 assumes that the probability is linearly dependent on temperature with the inclusion of an error term.

$$\pi = \beta_0 + \beta_1 \text{Temp}$$

This is already flawed because the model does not bound π in any way, whereas we know that probability is bounded between 0 and 1.

CLM 2 assumes random sampling. This runs into the same issues as discussed with logistic regression, which is that each O-ring in our example is not necessarily independent, especially O-rings of the same launch. However, we must make this assumption in both models the same way the authors did.

CLM 3: No perfect multi-collinearity

This is not an issue here since we only have a single feature. As long as there is any variation in the Temperature (which we do), this issue is satisfied.

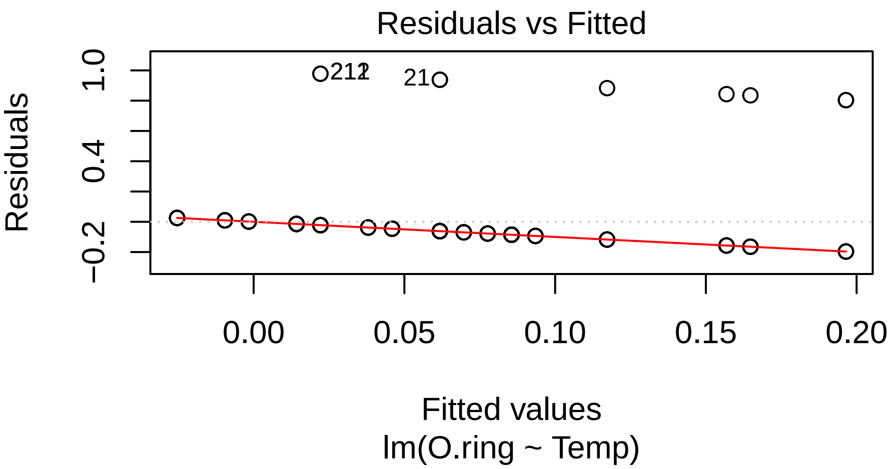
CLM 4: Zero Conditional Mean

This assumption states that the expectation of the error conditional on Temperature is 0.

$$E(u|\text{Temp}) = 0$$

Under zero conditional mean, we expect that the residuals on the residuals versus fitted value plot to have an expected value of 0 across the board. To check this, we plot the residual agains the fitted values for our set.

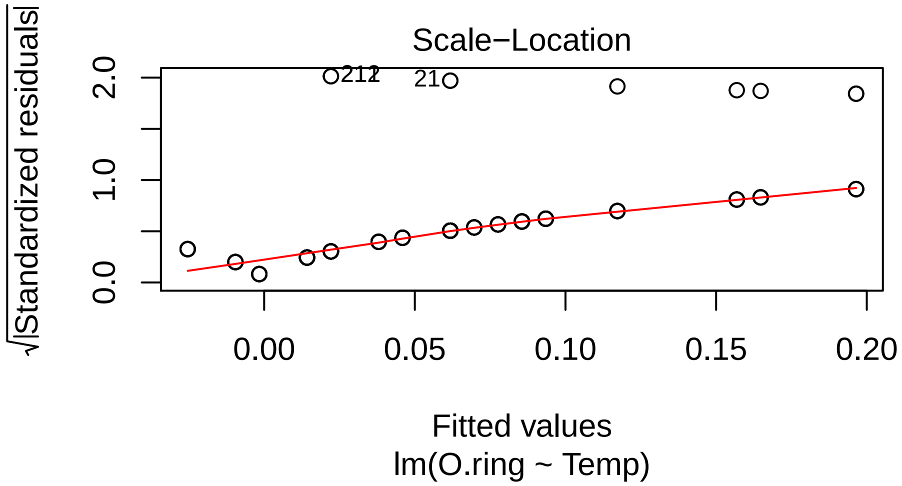
```r
plot(mod.lm, which = 1)
```



We see that there are two distinct lines, one for variables where the actual outcome of O-ring=0, and one for where O-ring=1. However, the mean of the residuals is certainly not 0 for all fitted values. Therefore, we fail to satisfy CLM 4.

CLM 5 assumes homoskedasticity, which is that the variance of the error terms are constant for all Temperature values. We know that homoskedasticity is violated because the outcome is binary, as the variance is $\pi * (1 - \pi)$. The easiest assessment is with the scale-location plot. If there is homoskedasticity, we would expect a roughly horizontal line of data points on this plot. This is clearly not the case below:

```
plot(mod.lm, which = 3)
```



One way to test for homoskedasticity is the Breusch-Pagan Test. The null hypothesis of the test states that we have homoskedasticity. We will test at a standard significance level of 0.05.

$$H_0 : \text{Homoskedasticity}$$
$$H_a : \text{Heteroskedasticity}$$

```
# lmtest was covered in 203 and heavily used for testing
# hypothesis surrounding the linear regression model.
# Therefore, I will use it here as well to test for
```

```r
# homoskedasticity.
library(lmtest)
```

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

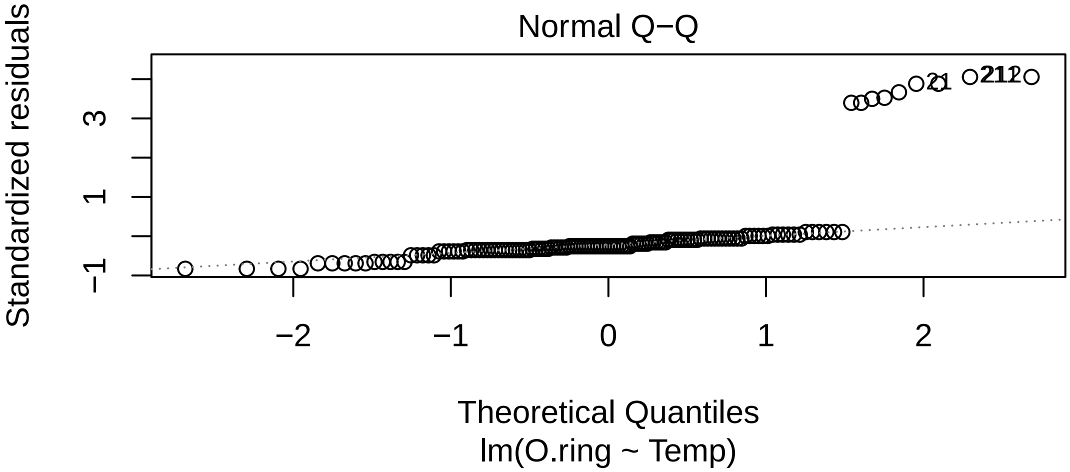
```r
bptest(mod.lm)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  mod.lm
## BP = 5.9172, df = 1, p-value = 0.01499
```

Since the $p - value < 0.05$, we reject the null hypothesis that we have homoskedasticity and have strong evidence for heteroskedasticity.

CLM 6 assumes that population error is independent of Temperature, and that the error term is normally distributed with mean 0 and constant variance. We can check this with the qqplot of the fitted values versus residuals plot.

```r
plot(mod.lm, which = 2)
```



We see that most of the points lie away from where we would expect them to be if the residuals are normally distributed. To test for normality, we can perform the Shapiro-Wilk test of normality.

We will test at a standard significance level of 0.05.

$$H_0 : \text{Residuals are normal}$$
$$H_a : \text{Residuals are not normal}$$

```r
shapiro.test(mod.lm$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  mod.lm$residuals
## W = 0.46777, p-value < 2.2e-16
```

Since p << 0.01, we reject the null hypothesis that the residuals are normal.

So in effect, CLM1,4,5,6 are all significantly violated. CLM 2 is likely violated in reality, but both models make this assumption.

Due to the fact many model assumptions are significantly violated, and the fact that probability is bounded between 0 and 1, whereas LPM has an unbounded support for the response, we greatly prefer logistic regression.

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

The main model is reproduced below without insignificant terms (Pressure and quadratic temperature):

$$\log \frac{\pi_{failure}}{1-\pi_{failure}} = 5.085 - 0.116 * \text{Temp}$$

One question is simply, during the day of the launch, which was 31 degrees, what are the odds and probability of failure at that particular temperature. The Odds of O-ring failure is:

$$\frac{\pi_{failure}}{1-\pi_{failure}} = e^{5.085 - 0.116 * \text{Temp}}$$

```
launch.Temp <- 31
exp(coef(mod.fit) %*% c(1, launch.Temp))[1]
```

```
## [1] 4.487703
```

The estimated odds of failure for launching at this particular day is 4.48. To look at the probability, this is:

$$\pi_{failure} = \frac{e^{5.085 - 0.116 * \text{Temp}}}{1 + e^{5.085 - 0.116 * \text{Temp}}}$$

```
exp(coef(mod.fit) %*% c(1, launch.Temp))[1]/(1 + exp(coef(mod.fit) %*%
    c(1, launch.Temp))[1])
```

```
## [1] 0.8177744
```

In other words, there was an estimated 0.818 chance that a particular O-ring would fail. Recall that each O-ring is assumed to be independent of each other, so if there are 6 rings, then $6 * 0.818 = 4.91$, then an expected 4.91 O-rings would fail, or almost all but 1 in the rocket. This means there would be a very high rate of incidence (failure for the launch itself) as well. In fact, the authors suggested that a single O-ring failure (binary model) would already be robust to number of incidents. To find the probability that at least one fails:

```
1 - (1 - 0.8177744)^6
```

```
## [1] 0.9999634
```

Based on this analysis, the probability that at least 1 O-ring fails is essentially 1, meaning the probability of an incident is essentially 1. We would definitely recommend postponing the launch based on this analysis.

Another question is, if the launch was postponed until the Temperature rose to 53 degrees (as proposed), what is the odds and probability of an O-ring failure then?

```
proposed.temp <- 53
odds <- exp(coef(mod.fit) %*% c(1, proposed.temp))[1]
prob <- exp(coef(mod.fit) %*% c(1, proposed.temp))[1]/(1 + exp(coef(mod.fit) %*%
```

```r
    c(1, proposed.temp))[1])
cbind(odds = odds, prob = prob)
```

```
##            odds      prob
## [1,] 0.3527892 0.2607865
```

The odds of an O-ring failure is 0.353, and the probability of O-ring failure 0.261, corresponding to an expected 1.57 O-ring failures. The probability of at least one failure is:

```r
1 - (1 - 0.2607865)^6
```

```
## [1] 0.8368379
```

This probability is still high, but it is better than the chances at 31 degrees. Postponing to even higher temperature than 53 would be recommended based on this model.

Another question is generally, how does Temperature affect the odds of O-ring failure. During the meeting, people were unsure whether Temperature has a non-negligble effect. It is reasonable to consider the change in odds of O-ring failure for a 10 degree increase in Temperature, since the temperature the day of launch was much less than previous test dates. We will give the estimate, as well as a confidence interval. The estimate is given by:

$$\begin{aligned} OR &= \frac{Odds_{T-10}}{Odds_T} \\ &= \frac{e^{5.085-0.116*(T-10))}}{e^{5.085-0.116*T}} \\ &= e^{-10*-0.116} \\ &= 3.19 \end{aligned}$$

This says that the odds of failure increase by 3.19 times for every 10 degree decrease in Temperature. The confidence interval for profile LR is:

```r
beta.ci <- confint(mod.fit, parm = "Temp", level = 0.95)
ci <- rev(exp(-10 * beta.ci)) %>% as.numeric()
output <- rbind(profile_LR_0.95 = cbind(lower = ci[1], upper = ci[2]))
output
```

```
##         lower    upper
## [1,] 1.277239 8.350003
```

With 95% confidence, the odds of O-ring failure change by 1.277 to 8.350 for a 10 degree decrease in Temperature. Since 1 is not included in this interval, we can conclude that 10 degree decrease in Temperature is statistically significant, and certainly has an effect O-ring failure rates. This would have provided evidence that could have potentially settled the discussion at the pre-Challenger teleconference.

Citation

1. Siddhartha R. Dalal, Edward B. Fowlkes & Bruce Hoadley (1989) Risk Analysis of the Space Shuttle: Pre-Challenger Prediction of Failure, Journal of the American Statistical Association, 84:408, 945-957, DOI: 10.1080/01621459.1989.10478858

*This work was done as part of the W271 - Statistical Methods for Discrete Response, Time Series, and Panel Data course under the U.C. Berkeley Master of Information and Data Science program. I thank Jessie Chen and Anu Yadav for their discussion and participation.