W271 Group Lab 1

Investigating the 1986 Space Shuttle Challenger Accident

Please fill in with your names.

Contents

Abstract

This report will, indeed, be abstract. No, instead, describe your goals your approach, and what you learn. In this study, we are investigating the effect of physical condition such as temperature and pressure have on the odds of O Rings failure on a space shuttle.

1 Introduction

1.1 Research question

How do temperature and pressure affect the odds of o-ring failures on a shuttle launch?

Do physical conditions, in term of temperature and pressure, affect the odds of O-Rings failures on a shuttle?

2 Data (20 points)

```
##      Flight          Temp          Pressure          O.ring          Number
##  Min.   : 1.0   Min.   :53.00   Min.   : 50.0   Min.   :0.0000   Min.   :6
##  1st Qu.: 6.5   1st Qu.:67.00   1st Qu.: 75.0   1st Qu.:0.0000   1st Qu.:6
##  Median :12.0   Median :70.00   Median :200.0   Median :0.0000   Median :6
##  Mean   :12.0   Mean   :69.57   Mean   :152.2   Mean   :0.3913   Mean   :6
##  3rd Qu.:17.5   3rd Qu.:75.00   3rd Qu.:200.0   3rd Qu.:1.0000   3rd Qu.:6
##  Max.   :23.0   Max.   :81.00   Max.   :200.0   Max.   :2.0000   Max.   :6

## 0 missing values in the data set.
```
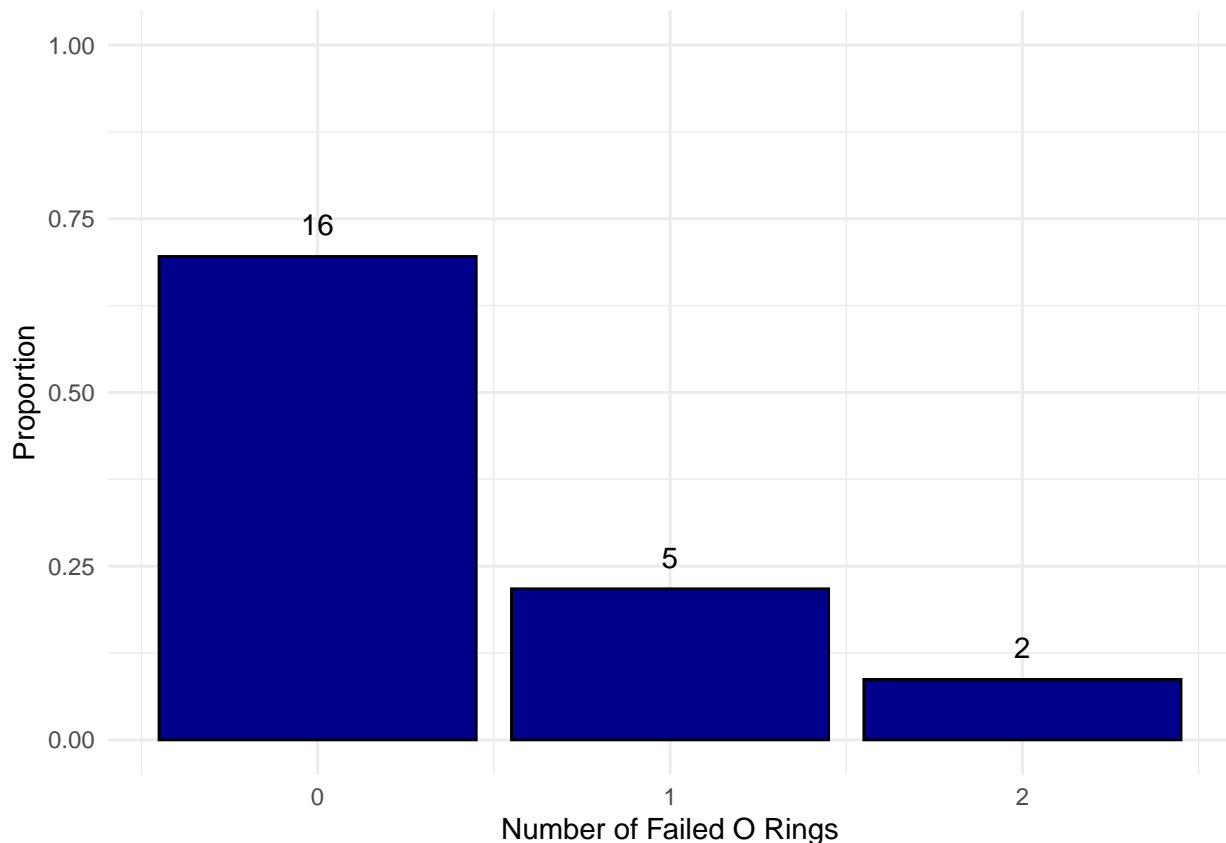
2.1 Description

The data for this analysis is from the 23 pre-Challenger launches of the space shuttle. It contains 5 columns: Flight, Temp, Pressure, O.ring, Number. Flight is a unique identifier representing each flight observed. Number represents the total number of O-rings in each spacecraft (6 for each spacecraft). O.ring is a random variable that denotes the number of O-rings failed at each launch. Lastly, Temp and Pressure are numerical values for the physical condition measurements taken at each launch.

There are 23 launches in the data and no missing values in any of the features in the data set. Temperature has 16 distinct values with a range between 53 to 81 and a mean of 69.57. Pressure has 3 distinct values with a range between 50 to 200 and a mean of 152. Both temperature and pressure records are full integer value. In terms of the response variable, number of failed O-rings, the majority of flights have 0 failed O-ring (16 out of 23). For the failed launch, there are 5 observations with 1 failed O-ring and 2 observations with 2 failed O-rings.

In the original study, the researchers treated each of the 6 O-rings as independent for each of the 23 launches. In other words, the failure of one O-ring do not cause any of the other O-ring to fail. This assumption is necessary to satisfy the binomial model's assumption of independence. The binomial model also assumes that each of the O-rings has equal probability to be damaged by either temperature or pressure. But when considering each O-ring as independent, the researcher might not account for the fact that a single failed O-ring could cause the other o-ring to fail and the entire launch to fail. While careful evaluation of the potential dependency between each O-ring is
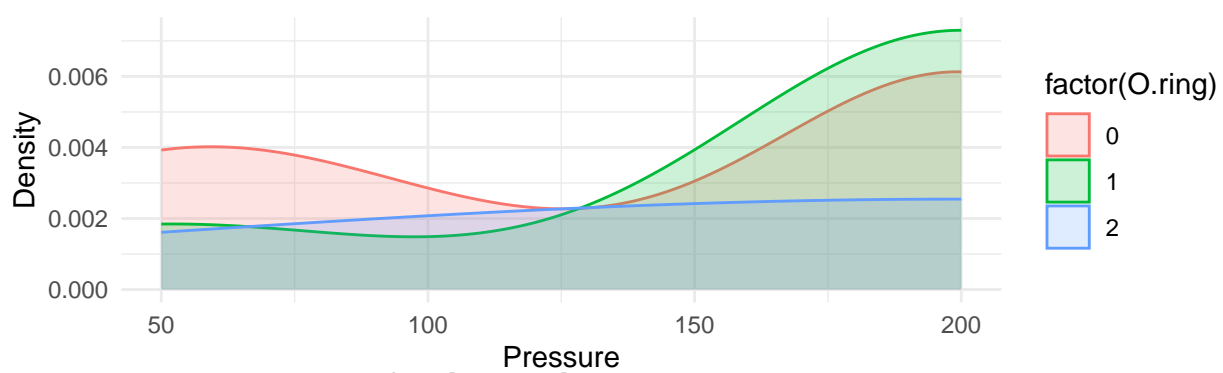
necessary, we will continue our analysis follow the independence assumption.
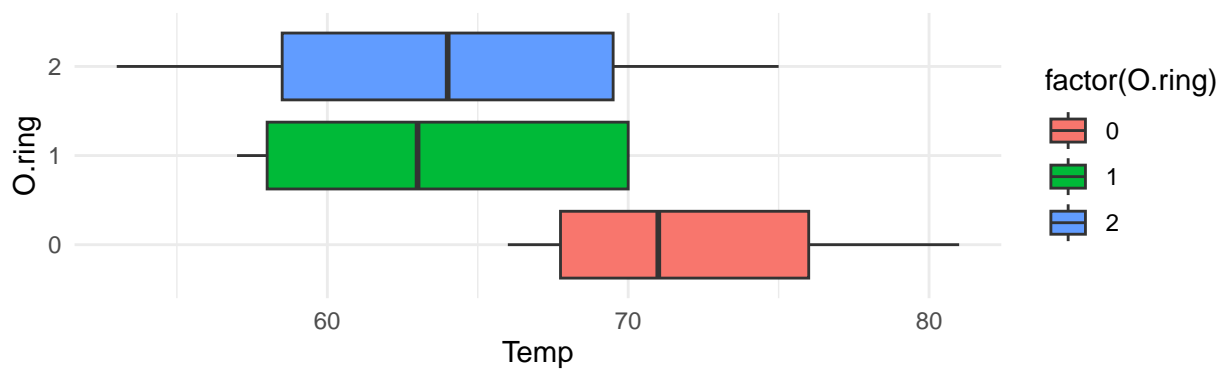
2.2 Key Features

Distribution of Temp
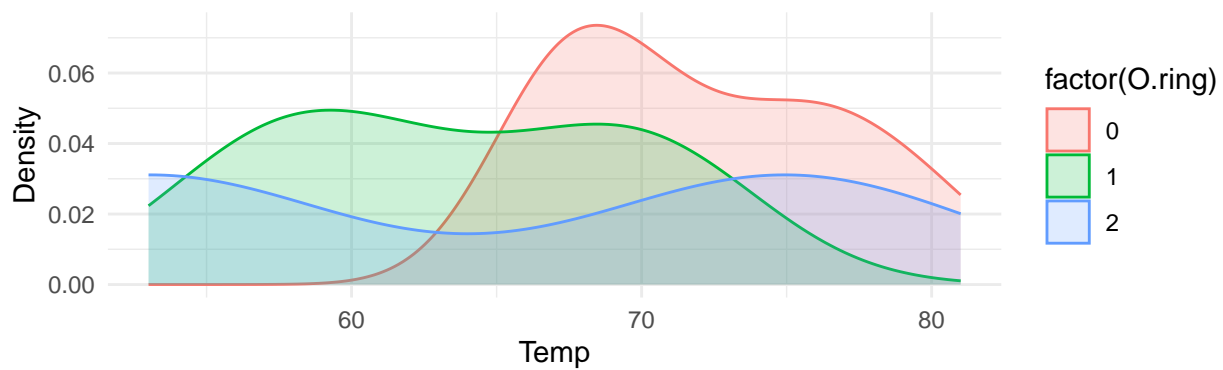
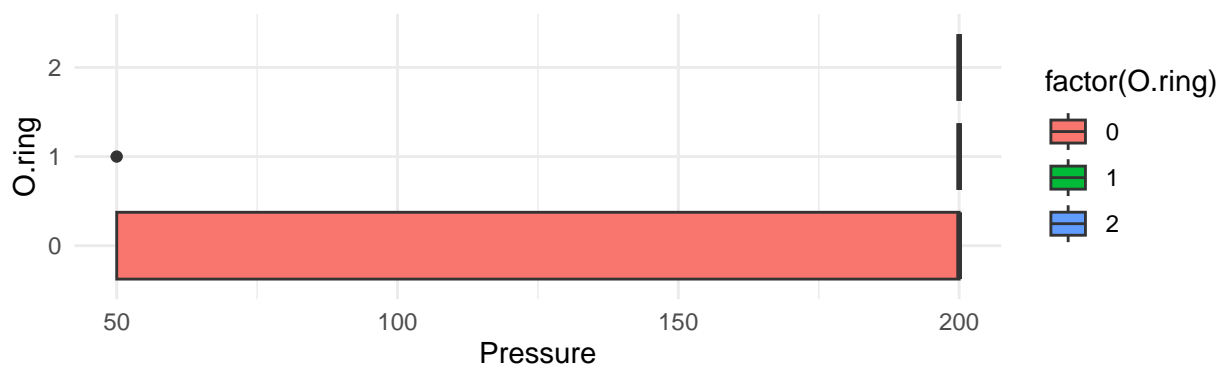Distribution of Pressure

Temp by Number of Oring Incidents

Pressure by Number of Oring Incidents

From the above study, we noted two features, temperature and pressure, that most likely affect the odds of O-rings failure on a space shuttle.From the distribution and box plot, the number of O-rings failed has a higher distribution at the lower temperature.

In term of pressure, the distribution plots showed that 0 failed O-rings at both high and low pressure while 2 failed O-rings showed a uniform distribution. However, it is noted that this plot does somewhat show that one O-rings failed has higher distribution in the higher pressure cases vs. the lower pressure cases. In the following study, we will be using both temperature and pressure as explanatory variable to model the odds of each individual O-ring fail during the launch. Our goal is to shed some light into the statistical significance of each of this feature on O-ring failure.

3 Analysis

3.1 Reproducing Previous Analysis (10 points)

Our baseline model is a binomial logistic regression model replicates what the authors present in their report.

```
# Since there are 6 O-rings on a shuttle We define the failure rates as number
# of O Rings Failed out of 6 O Rings create a binomial logistic regression
# model to evaluate the odds of O ring failure
mod1 <- glm(formula = cbind(O.ring, Number - O.ring) ~ Temp + Pressure, family = binomial(link
    data = challenger)
```

```
##
## Table 1: The Estimated Relationship Between O-Rings Failure and Temperature/Pressure
## ============================================================
##                     Output Variable: Log Odds of O-Rings Failure
##                     ----------------------------------------------
##
## ------------------------------------------------------------
## Temperature                        -0.098*
##                                    p = 0.029
##
## Pressure                           0.008
##                                    p = 0.270
##
## (Intercept)                        2.520
##                                    p = 0.470
##
## ------------------------------------------------------------
## Observations                       23
## Log Likelihood                     -15.053
## Akaike Inf. Crit.                  36.106
## ============================================================
## Note:                       *p<0.05; **p<0.01; ***p<0.001
```

```
# Odds ratio for model 1
OR <- round(exp(10 * mod1$coefficients), 3)[2]
```

For the first model, we perform a logistic regression on the odds of O-ring failed response using both temperature and pressure.

The estimated model is $y = 2.520 - 0.98 * Temp + 0.008 * Pressure$.

Using a Wald test, the model return a p-value of 0.029 for temperature and 0.270 for pressure. Since the p-value for Temperature is below the 95% confident level, Temperature is statistical significant to estimate the O-Rings failure. The p-value for pressure is above this level and is considered not statistically significant. Additionally, from the fitted model, we see that temperature have a coefficient of -0.098, indicating that an increase in temperature decrease the log odd of an O-ring failure. While pressure is not statistically significant, the positive coefficients suggested that an increase in pressure increase the log odd of an O-ring failure. These observations are aligned with what we saw from the distribution and box plot charts in our EDA analysis. We took this analysis further by computing the Odd Ratios for each estimated coefficients for a 10 units increase in temperature or pressure. Here, the odds of an O-ring failure is 0.374 times as likely, or about 63% less likely, for a 10 unit increase in temperature. The pressure yield an Odd Ratio of 1, which indicate that an increase in pressure have similar effect on the odds of an O-ring failure as no increase in pressure.

```
Anova(mod1, test = "LR")
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: cbind(O.ring, Number - O.ring)
##           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
```

In order to test for the significance of Pressure, we use the Likelihood Ratio Test for each explanatory variable. For this test, we stated a Null hypothesis that $\beta_2 = 0$, or Pressure does not add significant explanatory power for the odds of O-ring failure, on top of the Temp in the model.

From the test, we obtained a p-value for a Chi-squared distribution of 0.0228 for Temperature and 0.2145 for pressure. Like the Wald test, the p-value for Pressure is above the 95% confident level, alpha=0.05, so there is not enough evidence to reject the Null hypothesis.

Considering both test results, we conclude that Pressure does not add much explanatory power to the model. Excluding it from the model is especially recommended in this case, considering the sample size of 23 observations, the drop in the degree of freedom is 5% when we add one additional explanatory variable. AIC score also decreases with one fewer variable and lower AIC is better. Keeping the model simple and easy to interpret would also help with the further study.

3.2 Confidence Intervals (20 points)

After defining Temperature as our explanatory variable in the previous section, an anova test determined whether a quadratic term is needed in the logistic regression model.

```
# Estimate the logistic regression model
mod2 <- glm(formula = cbind(O.ring, Number - O.ring) ~ Temp, family = binomial(link = "logit")
    data = challenger)
mod3 <- glm(formula = cbind(O.ring, Number - O.ring) ~ Temp + I(Temp^2), family = binomial(link
    data = challenger)
```

```
##
## Table 2: The Estimated Relationship Between O-Rings Failure and Temperature Effects
## ================================================================
##                      Output Variable: Log Odds of O-Rings Failure
##                      ------------------------------------------------
##
##                              (1)                        (2)
## ----------------------------------------------------------------
## Temperature                 -0.116*                    -0.651
##                           p = 0.014                  p = 0.380
##
## Squared Temperature                                    0.004
##                                                      p = 0.467
##
## (Intercept)                  5.085                     22.126
##                           p = 0.096                  p = 0.353
##
## ----------------------------------------------------------------
## Observations                  23                        23
## Log Likelihood             -15.823                    -15.576
## Akaike Inf. Crit.           35.647                     37.152
## ================================================================
## Note:                                    *p<0.05; **p<0.01; ***p<0.001
```

```
anova(mod2, mod3, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: cbind(O.ring, Number - O.ring) ~ Temp
## Model 2: cbind(O.ring, Number - O.ring) ~ 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
```
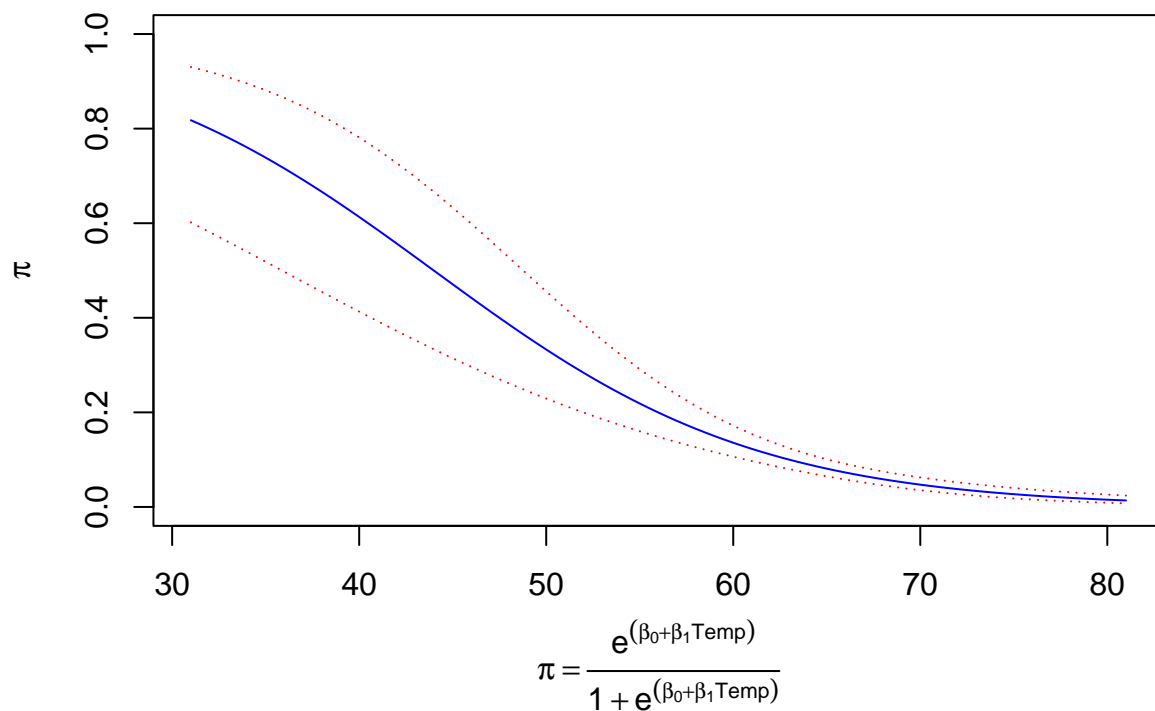
The anova test yielded a p-value of 0.48, which means that there is little to no evidence that a quadratic term is statistically significant. Therefore we fail to reject the null hypothesis that mod3 is equal to mod2. Therefore, we will move forward with mod3.
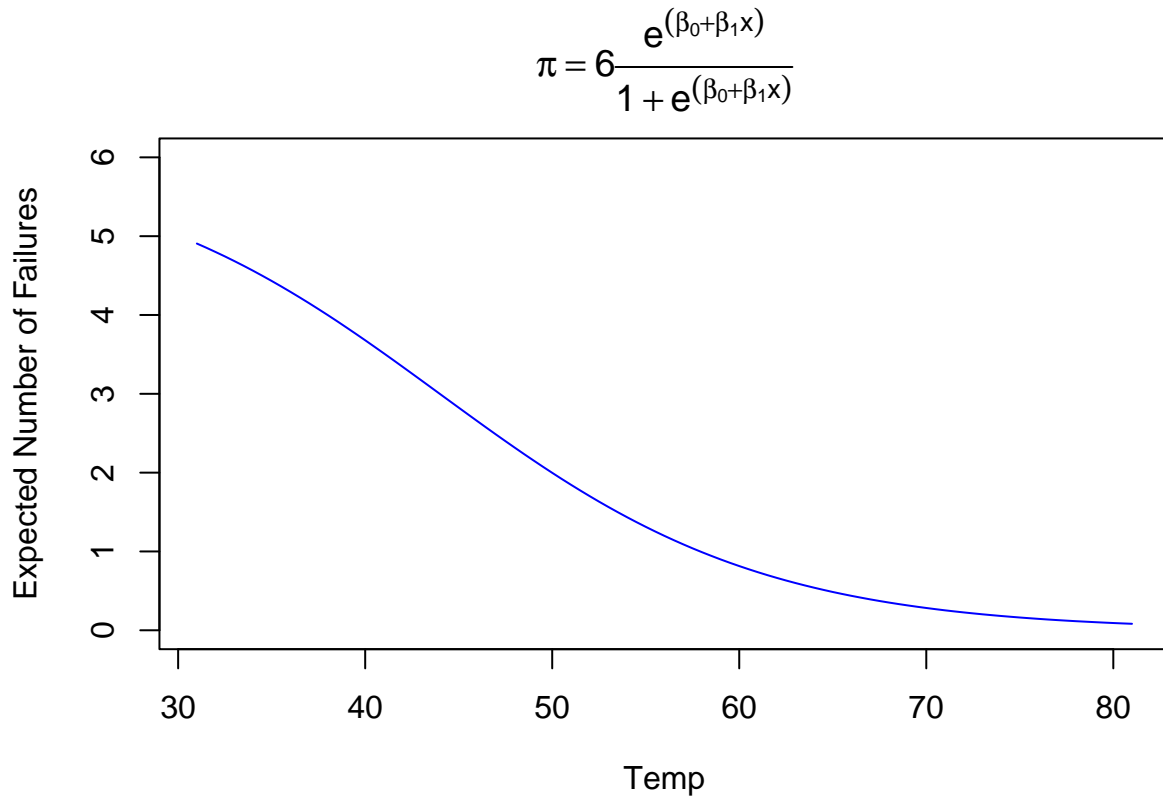
```
mod.beta <- mod2$coefficients
# calculate true wald confidence interval at temperature
wald.CI.true.coverage <- function(mod, x_value) {
    alpha <- 0.5
    predict.data <- data.frame(Temp = x_value)
    linear.pred <- predict(object = mod, newdata = predict.data, type = "link", se = TRUE)
    CI.lin.pred <- linear.pred$fit + qnorm(p = c(alpha/2, 1 - alpha/2)) * linear.pred$se.fit
    CI.pi <- exp(CI.lin.pred)/(1 + exp(CI.lin.pred))
    wald.df <- round(data.frame(x_value, lower = CI.pi[1], upper = CI.pi[2]), 4)
    return(wald.df)
}
x_seq <- seq(31, 81, by = 1)
wald.CI.true.matrix <- matrix(data = NA, nrow = length(x_seq), ncol = 3)
counter <- 1
for (x_value in x_seq) {
    wald.df2 <- wald.CI.true.coverage(mod2, x_value)
    wald.CI.true.matrix[counter, ] <- c(x_value, wald.df2$lower, wald.df2$upper)
    counter <- counter + 1
}
# check that the matrix is correct wald.CI.true.matrix[1:5, ]
```

Probability of O−ring Failure vs. Temp

$$\pi = \frac{e^{(\beta_0 + \beta_1 \text{Temp})}}{1 + e^{(\beta_0 + \beta_1 \text{Temp})}}$$

$$\pi = 6\frac{e^{(\beta_0 + \beta_1 x)}}{1 + e^{(\beta_0 + \beta_1 x)}}$$

From the above chart, odds of O-Rings failure declines as temperature increases, as β_1 is negative. The confidence intervals are wider for temperatures lower than 50F and narrower for temperatures greater than 65F. The estimated interval for temperature equal to 30F is about (1,6); this illustrates that we have high variability for the expected number of incidents. The reason for wider confidence interval for lower than 50F is because the minimum temperature in the data set was 53F. The forecast error for the below-50F is larger.

```
# Compute the Probability at T=31 degree
params <- dim(summary(mod2)$coefficients)[1]
predict.data <- matrix(data = c(1, 31), nrow = 1, ncol = params)
linear.combo <- mcprofile(object = mod2, CM = predict.data)
mroz_logit_profile <- confint(linear.combo, level = 0.95)
mroz_logit_ci <- exp(mroz_logit_profile$confint)/(1 + exp(mroz_logit_profile$confint))
mroz_logit_estimate <- exp(mroz_logit_profile$estimate)/(1 + exp(mroz_logit_profile$estimate))
result.profile_lr <- data.frame(estimate = mroz_logit_estimate, lower = mroz_logit_ci$lower,
    upper = mroz_logit_ci$upper)
result.profile_lr
```

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

To apply the inference procedures, we assume asympotic properties of a large sample size, independence of each O-Ring failure and each failure has equal probability. At 31F, the model estimes 81.8% probability of O-Rings failure, with 95% confidence interval of 14.2% and 99%.

3.3 Bootstrap Confidence Intervals (30 points)

```r
# TODO: create a for loop to do this, for each data set: for each model from
# the resulting data set: compute the probability at each integer temeprature
# value (10 to 100) compute the CI for the probability
temp <- NULL
pi.hat <- NULL
CI.lower <- NULL
CI.upper <- NULL
for (i in 1:100) {
    challenger_resampled <- NULL
    # creating a resampled data from the original dataset with replacement
    challenger_resampled <- challenger[sample(nrow(challenger), size = 23, replace = TRUE),
        ]
    # construct a model from the new resampled dataset
    mod_resample <- glm(formula = cbind(O.ring, Number - O.ring) ~ Temp, family = binomial(link
        data = challenger_resampled)
    output <- NULL
    conf.lower <- NULL
    conf.upper <- NULL
    tprime <- NULL
    # run the model though to obtain the output
    for (t in 10:100) {
        new.data <- data.frame(Temp = t)
        pred.prob <- predict(mod_resample, newdata = new.data, type = "link", se = TRUE)
        # compute probability and CI
        pi <- exp(pred.prob$fit)/(1 + exp(pred.prob$fit))
        linear.Wald.CI <- pred.prob$fit + qnorm(p = c(0.05, 0.95)) * pred.prob$se
        Wald.CI <- exp(linear.Wald.CI)/(1 + exp(linear.Wald.CI))
        # append data to array
        tprime <- c(tprime, t)
        output <- c(output, pi)
        conf.lower <- c(conf.lower, Wald.CI[1])
        conf.upper <- c(conf.upper, Wald.CI[2])
    }
    # append each bootstrap run
    temp <- c(temp, tprime)
    pi.hat <- c(pi.hat, output)
    CI.lower <- c(CI.lower, conf.lower)
    CI.upper <- c(CI.upper, conf.upper)
}
bootstrap_ouptuts <- data.frame(temp = temp, pi.hat = pi.hat, lower = CI.lower, upper = CI.upp
agg_tbl <- bootstrap_ouptuts %>%
    group_by(temp) %>%
    summarise(pi.hat = mean(pi.hat), lower = mean(lower), upper = mean(upper))
```
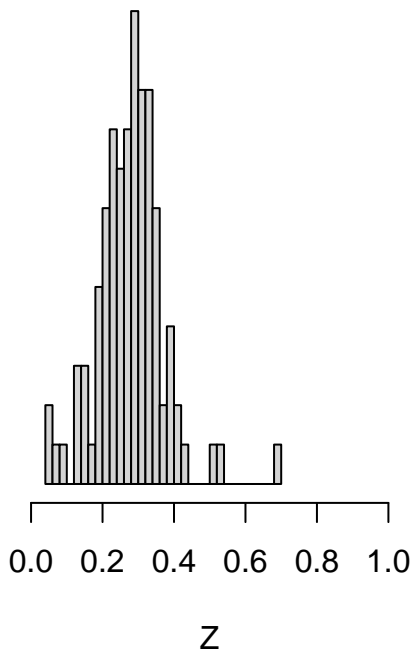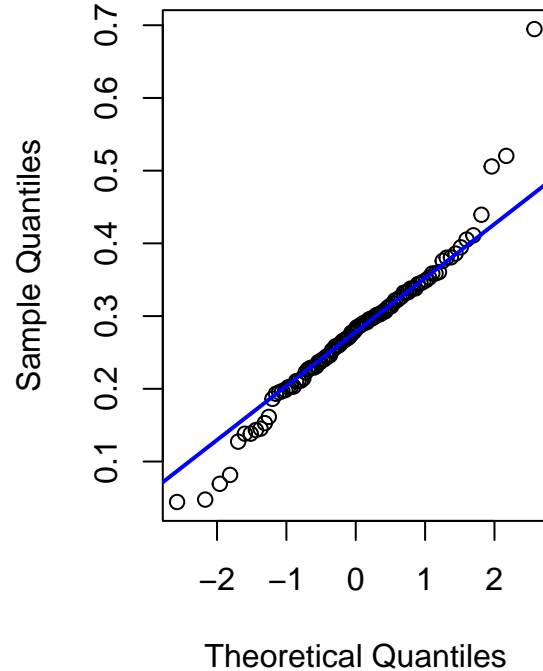
T = 31 Gaussian Distribution
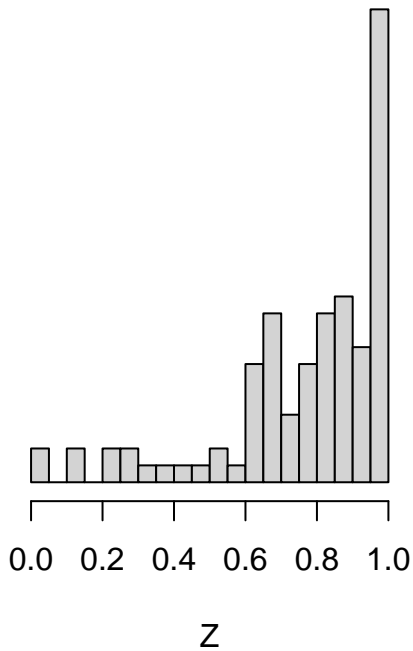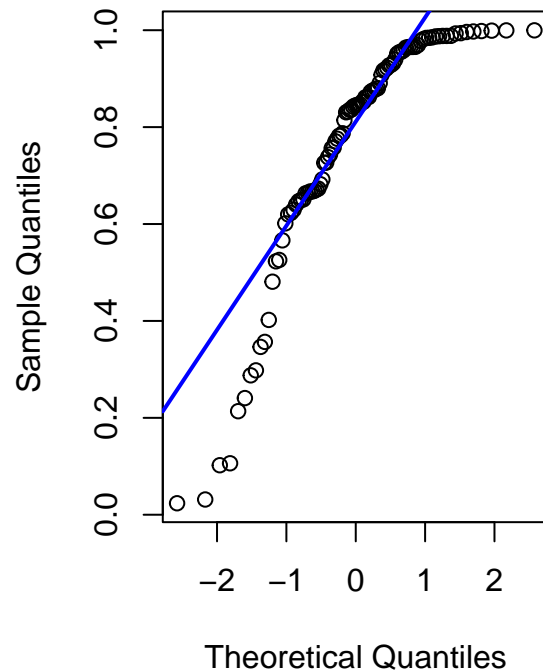
Normal Q–Q Plot

T = 53 Gaussian Distribution

Normal Q–Q Plot

Statistical bootstrapping is a common method use when we want to generate some statistics with reasonable confidence. Since we are not able to perform launching experiment for large N, bootstrapping allows us to simulation such a process. In this study, we performed 100-folds bootstrap

of the challenger datataset so that we can obtain 100 probability of failure at each temperature ranging from 10° to 100° Fahrenheit. It is worth nothing here that at each iteration, we're sampling 23 observations from the original challenger dataset with replacement so that we do not have the same set of values in each iteration. After the dataset is sampled, we used the new re-sampled dataset to fit a model for the log Odd of O-ring Failure using only temperature as explanatory variable. By fitting a new model for each re-sampled dataset, we are able to obtain slightly different coefficient that would ultimately yield a slightly different prediction at each iteration. Finally, we record the probability of an O-ring failure as well as the 90% Wald confidence interval limits given an integer temperature between 10° to 100° Fahrenheit using each resampled model.

From the result obtained, we sliced the data on two temperatures values 31° F and 53° F to reference the temperature at which the Challenger spaceship launch failed and the lowest launch temperature. From the two sliced data, we generated a histogram of our 100 bootstrap sampled as well as the Q-Q plot to illustrate the predicted probability deviations versus the theoretical Gaussian distribution. Here, we are able to see that due to having literally no launches at 31° F, the failure probability distribution is sparse but shown to be aggregate at high probability region between 0.7-1.0. In contrast to 53° F, the failure probability distribution is more Gaussian-like and aggregate at low probability region between 0.0-0.4. The Q-Q plots for both temperatures values further support this by showing the predicted failure probability versus the theoretical failure probability at each quantiles. Noted that the 53° F Q-Q plot shows an almost 1-1 match between the two quantiled probability while there is a clear deviation near the tail ends for the 31° F Q-Q plot.

While it is safe to say that if we're able to replicate the launch measurement at each temperature values 100 more times, we'll have a better idea of the probability of failure at each temperature. Using the bootstrapping method, we're able to somewhat show the effect of launch temperature of odds of an O-ring failure. In short, the temperature at which the Challenger space craft launch at, 31° F has a really high probability of O-ring failure. Delaying the launch to a warmer temperature would have helped the outcomes

3.4 Alternative Specification (10 points)

```
model.linear <- lm(formula = I(O.ring/Number) ~ Temp, data = challenger)
```

```
##
## Table 3: Linear Effect of O-Rings Failure and Temperature
## ==================================================
##                          Dependent variable:
##                      -----------------------------
##                           I(O.ring/Number)
## --------------------------------------------------
## Temp                           -0.008*
##                               p = 0.013
##
## Constant                       0.616**
##                               p = 0.007
##
## --------------------------------------------------
## Observations                     23
```

```
## R2                          0.261
## Adjusted R2                 0.226
## Residual Std. Error      0.096 (df = 21)
## ==================================================
## Note:                   *p<0.05; **p<0.01; ***p<0.001
```

Since pressure was not statistically significant, we only use temperature as an explanatory variable to a linear regression to model the probability of failure, defined as number of O-rings failed divided by the total number of O-rings. With this model, we obtained a p-value of 0.013 for temperature, which is lowered than the stated 95% Wald confidence level, or alpha 0.05 and visualized by 1 asterisk. The negative coefficient for temperature suggests that at an increase in pressure decreases the probability of failure.

While that is all well and good, performing a linear regression in this aspect violates the model identically independent distribution assumption. The given dataset has the response variable O.ring as an integer count from O to 6. And the original research treated each O.ring as an independent observation. Thus, by simply dividing O.ring by the maximum number of O.ring and performing a linear regression, we are sometime bundling more than one failled O-rings together in one observation. In this case, if we wanted to model this properly based on the i.i.d. assumptions, a logistic regression would have to be use.

4   Conclusions (10 points)

```
summary(mod2)
```

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

```
## Number of Fisher Scoring iterations: 5
```

```
round(exp(coef(mod2)), 4)
```

```
## (Intercept)        Temp
##     161.5763      0.8908
```

In conclusion, our final model is a binomial logistic regression model with o-ring success as the dependent variable and temperature as the explanatory variable. The table above is the odds ratio of having a 1 unit increase from that corresponding temperature. For example, the odds of an o-ring failing increases by 0.8908 times for every 1-unit increase in temperature.

Our explanatory variable analysis involved anova tests comparing models with combinations of Temperature, Pressure, and Temperature-squared. The results demonstrate that Model 3, with Temperature as the only explanatory variable for O.Ring failure, is the most statistically significant model. In this model, the probability of o-ring failure decreases as temperature increases, and the confidence interval also dramatically decreases when the temperature is greater than 65 degrees compared to less than 65 degrees. This observation is reinforced through parametric bootstrapping. With 100 iterations of bootstrapping and 23 observations per iteration, the results showed that at the lowest temperature of 31F, the failure probability distribution is sparse but aggregates at a high probability region between 0.7 and 1.0 while at 53F, the failure probability distribution is Gaussian-like and aggregates at a low probability region between 0.0 and 0.4. Therefore, delaying the launch to a warmer temperature decreases the likelihood of o-ring failure. Finally, we considered a linear regression model as an alternative specifications against the binomial model, but doing so would violate the assumption of an identically independent distribution.