

W271 Lab 1

Tiffany Jaya, Joanna Huang, Shan He, Robert Deng

```
# add packages
library(dplyr)
library(ggplot2)
library(knitr)
library(stargazer)
# prevent source code from running off the page
opts_chunk$set(tidy.opts=list(width.cutoff=70), tidy=TRUE)
# remove all objects from current workspace
rm(list = ls())
# set seed number to reproduce results
set.seed(101)
# load data
d <- read.csv("../dataset/challenger.csv", header=TRUE, sep=",")
# set x-axis min and max for temp, temp at launch, recommended temp to launch
xlim.min <- 31; xlim.max <- 81; temp.launch <- 31; rec.temp <- 72
```

Introduction

On January 28, 1986, the space shuttle Challenger exploded after a mere 73 seconds into its flight, resulting in the death of all seven crew members. An investigation ensued and found that the accident was caused by the O-rings failing to seal the joints properly in the rocket booster. Two potential variables were investigated: temperature and pressure. Based on findings at the time, researchers surmised that when the temperature is cold, similar to the temperature at the time of the launch at 31°F, O-ring retains its cold-compressed shape, causing the pressurized gas from within to leak out and initiate the disintegration of the vehicle. Furthermore, there was a change in the leak check procedure from 50 psi to 200 psi to check the position of the O-ring.

Based on this background information, we seek to retrospectively analyze and predict how the specific conditions of temperature and pressure may have led to the Challenger explosion. Using logistic regression modelling, we answer the question: Were the temperature at the time of the launch or the change in the leak check procedure contributing factors to the O-rings failure, and consequently the structural breakdown of the space shuttle Challenger?

EDA

In order for us to have a better understanding of this relationship, we explored the dataset that contains 23 observations, one for each of the prior flights Challenger took before its final one. Looking at the dataset, there are five numerical attributes with no missing values:

- Flight: the temporal order of flights
- Temp: the temperature at the time of launch (in degree Fahrenheit)

- Pressure: the air pressure used in the leak check procedure (in psi)
- O.ring: the number of O-rings failure on the given flight
- Number: the total number of O-rings, which remains constant at six total, because the rocket booster has three field joints with two O-rings each. This variable has no bearing on O-rings failure.

```
# structure of data
str(d)
```

```
## 'data.frame':   23 obs. of  5 variables:
## $ Flight   : int  1 2 3 4 5 6 7 8 9 10 ...
## $ Temp     : int  66 70 69 68 67 72 73 70 57 63 ...
## $ Pressure: int  50 50 50 50 50 50 100 100 200 200 ...
## $ O.ring   : int  0 1 0 0 0 0 0 0 1 1 ...
## $ Number   : int  6 6 6 6 6 6 6 6 6 6 ...
```

```
# summary of data
kable(summary(d))
```

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

```
# list row with missing values
d[!complete.cases(d), ]
```

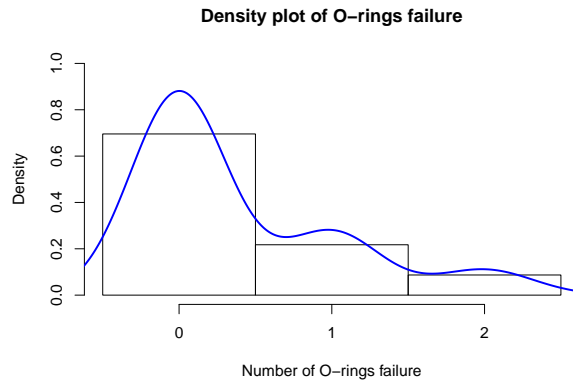
```
## [1] Flight Temp Pressure O.ring Number
## <0 rows> (or 0-length row.names)
```

Univariate analysis of O.ring

```
# frequency of O-rings failure
table(d$O.ring)
```

```
##
##  0  1  2
## 16  5  2
```

```
# density of O-rings failure
hist(d$O.ring, xlab = "Number of O-rings failure", main = "Density plot of O-rings failure",
      breaks = 0:3 - 0.5, xaxt = "n", ylim = c(0, 1), freq = FALSE)
axis(1, at = seq(min(d$O.ring), max(d$O.ring)))
lines(density(d$O.ring), col = "blue", lwd = 2)
```

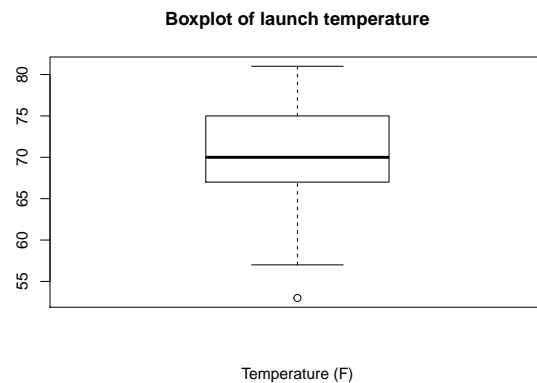
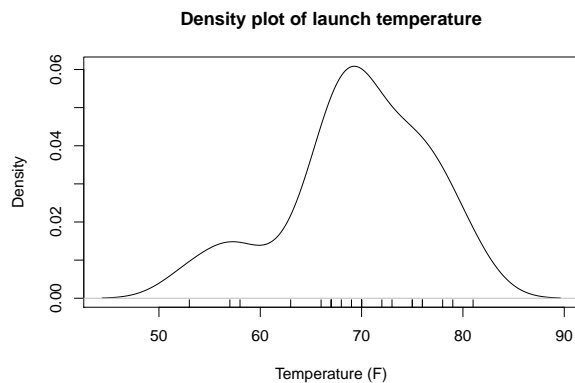


Since the disaster occurred as a result of O-ring failure, the O.ring variable will be the variable that we are trying to predict. From the density plot of the sample, we can approximate that an event of no O-ring failure is more than twice as likely to occur than a single O-ring failure and seven times more likely to occur than a second O-ring failure.

Univariate analysis of Temp

```
# density of launch temperature
plot(density(d$Temp), main = "Density plot of launch temperature", xlab = "Temperature (F)")
rug(d$Temp)
# boxplot of launch temperature
boxplot(d$Temp, main = "Boxplot of launch temperature", xlab = "Temperature (F)")
# display temperature outlier
d %>% filter(Temp == min(Temp))
```

```
## Flight Temp Pressure O.ring Number
## 1 14 53 200 2 6
```



As a potential cause of the O-ring failure, Temp is another variable that we analyze. Its density distribution is slightly skewed to the left with the mean being smaller than the median (see summary table), meaning they perform more test launches when the temperature is warmer. It is also interesting to note that two O-rings fail at the outlier temperature of 53 °F and the updated leak check pressure of 200 psi.

Univariate analysis of Pressure

```

# frequency of air pressure used in leak check procedure
table(d$Pressure)

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

# the leak check procedure used in flights
sapply(unique(d$Pressure), function(p) c(min(which(d$Pressure %in% p)),
max(which(d$Pressure %in% p))))

##      [,1] [,2] [,3]
## [1,]    1    7    9
## [2,]    6    8   23

```

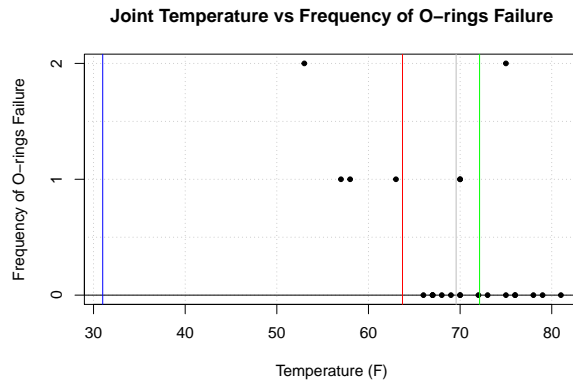
In addition to Temp, we analyze Pressure, the second potential predictor of O-rings failure. The leak check procedure uses air pressure to check the position of the O-ring and changes 3 times: flight 1-6 using an air pressure of 50 psi, flight 7-8 using 100 psi, and flight 9-23 using 200 psi. The increase in air pressure for the leak check procedure is mostly due to a growing desire to ensure that a leak is not masked. Later in the report, we will analyze if the increase in air pressure corresponded to an increase in the number of O-rings failing.

Bivariate analysis

```

# average temperature for flights with at least one O-rings failure
avg.temp.failure <- d %>% filter(., O.ring > 0) %>% .$Temp %>% mean(.)
# average temperature for flights with no O-rings failure
avg.temp.success <- d %>% filter(., O.ring == 0) %>% .$Temp %>% mean(.)
# average temperature for all flights
avg.temp <- mean(d$Temp)
# plot joint temperature vs number of O-rings failure incidents
plot(main = "Joint Temperature vs Frequency of O-rings Failure", x = d$Temp,
      xlab = "Temperature (F)", xlim = c(xlim.min, xlim.max), y = d$O.ring,
      ylab = "Frequency of O-rings Failure", yaxt = "n", panel.first = grid(col = "gray",
      lty = "dotted"), pch = 20)
abline(h = 0, v = c(avg.temp.failure, avg.temp.success, avg.temp, temp.launch),
      col = c("black", "red", "green", "gray", "blue"))
axis(side = 2, at = 0:2)

```



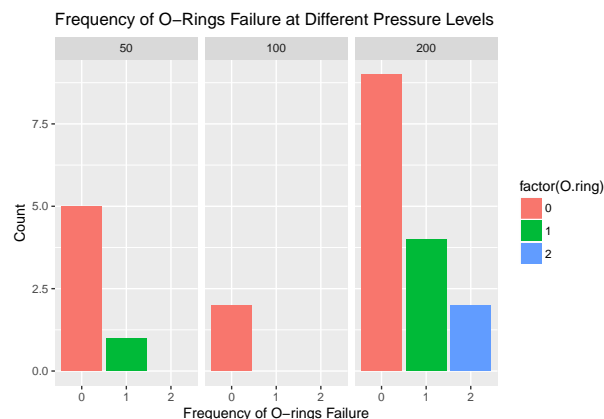
When we performed a bivariate analysis of Temp and O.ring variables, we started to notice that if we only looked at the average temperature of flights with at least one O-ring failure (red line), the number of O-rings failing in cold weather is approximately similar to that in the warm weather. It was these specific data points used by two staff members as a reasoning to launch Challenger at 31°F temperature (blue line). But if we took into consideration all previous flights, the average temperature of all 23 launches (gray line) or the average temperature of all successful launches (green line) give us clue that O-rings seem to perform better in warmer temperature.

```
summary(aov(Temp ~ O.ring, data = d))
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## O.ring        1  286.2   286.24    7.426  0.0127 *
## Residuals    21  809.4    38.54
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A one-way anova test confirms our intuition that the mean temperature of successful launches is different than the mean temperature of failure launches with a p-value of 0.01 . This provides strong evidence to include temperature in the modeling section.

```
d %>% ggplot(aes(x = factor(O.ring), fill = factor(O.ring))) + geom_bar() +
  facet_wrap(~Pressure) + labs(title = "Frequency of O-Rings Failure at Different Pressure Levels",
  y = "Count", x = "Frequency of O-rings Failure")
```



Plotting with facets reveals that most O.ring failures occurred at 200 psi. The uneven levels of

pressure makes this an unattractive explanatory variable to model. To investigate pressures between failed and successful O-ring states, a one-way anova test reveals a p-value of 0.188 in which we fail to reject the null hypothesis of no difference between the 3 pressure states.

```
summary(aov(Pressure ~ O.ring, data = d))
```

##		Df	Sum Sq	Mean Sq	F value	Pr(>F)
##	O.ring	1	8297	8297	1.852	0.188
##	Residuals	21	94094	4481		