Module 9: Logistic Regression

Rebecca C. Steorts

Agenda

- 1986 Challenger explosion
- What happened?
- Background
- Logistic regression

Motivation

- In 1986, the Challenger space shuttle exploded as it took off.
- The question of interest was what happened and could it have been prevented?
- To understand this, we need to learn about to concepts:
- “o-rings" and Bayesian logistic regression

What is an “o-ring"

“An o-ring, also known as a packing or a toric joint, is a mechanical gasket in the shape of a torus. It is a loop of elastomer with a round cross-section, designed to be seated in a groove and compressed during assembly between two or more parts, creating a seal at the interface."

-Wikipedia

What is an “o-ring"

- O-rings are on component of a space shuttle, and we now know that they can fail at low temperatures.
- We did not know this in 1986 during the Challenger launch.

O-ring Data

```r
library(faraway)
data("orings")
orings[1,] <- c(53,1)
head(orings)
```

```
##   temp damage
## 1   53      1
## 2   57      1
## 3   58      1
## 4   63      1
## 5   66      0
## 6   67      0
```

O-ring Data

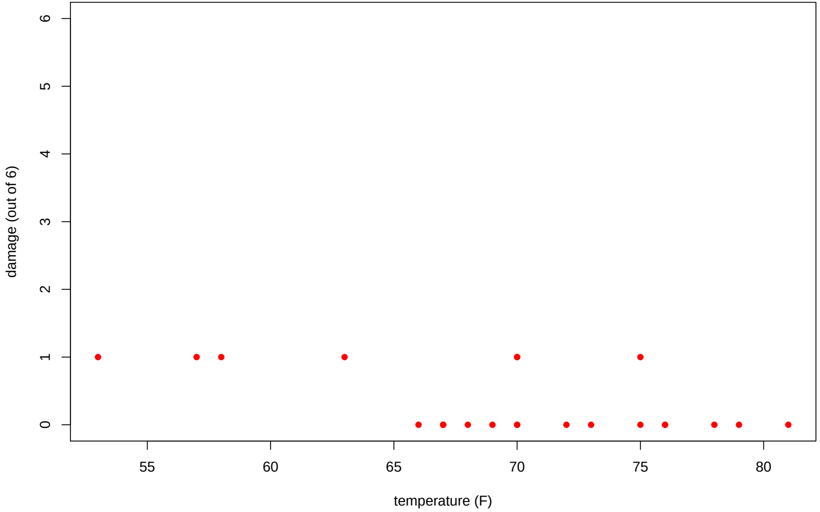
The 1986 crash of the space shuttle Challenger was linked to failure of o-ring seals in the rocket engines.

Data was collected on the 23 previous shuttle missions, where the following variables were collected:

- temperate for each mission
- damage to the number of o-rings (out of a total of six)

Plot

```
plot(damage~temp, data=orings, xlab="temperature (F)",
     ylab="damage (out of 6)",
     pch=16, col="red", ylim=c(0,6))
```



Plot

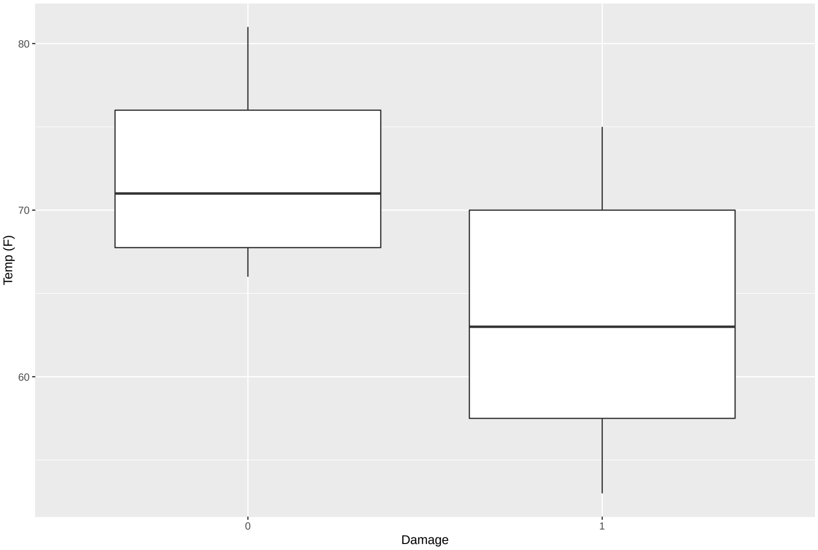
```r
library(ggplot2)
geom_boxplot(outlier.colour="black", outlier.shape=14,
             outlier.size=2, notch=FALSE)
```

```
## geom_boxplot: outlier.colour = black, outlier.fill = NUl
## stat_boxplot: na.rm = FALSE, orientation = NA
## position_dodge2
```

```r
damage <- as.factor(orings$damage)
temp <- orings$temp
head(damage)
```

```
## [1] 1 1 1 1 0 0
## Levels: 0 1
```

Boxplot of temperature versus o-ring failure



Response and covariate

- The response is the damage to the o-ring (in each shuttle launch).
- The covariate is the temperature (F) in each shuttle launch.

The question of interest is the probability of an o-ring failure.

Notation and Setup

- Let p_i be the probability that o-ring i fails.
- The corresponding **odds of failure** are

$$\frac{p_i}{1 - p_i}.$$

Notation and Setup

- The probability of failure p_i is between $[0, 1]$
- The odds of failure is any real number.

Logistic Regression

The response

$$Y_i \mid p_i \sim \text{Bernoulli}(p_i) \tag{1}$$

for $i = 1, \ldots, n$.

The logistic regression model writes that the logit of the probability p_i is a linear function of the predictor variable(s) x_i:

$$\text{logit}(p_i) := \log(\frac{p_i}{1 - p_i}) = \beta_0 + \beta_1 x_i. \tag{2}$$

Interpretation of Co-efficients

- The regression co-efficients β_0, β_1 are directly related to the log odds $log(\frac{p_i}{1-p_i})$ and not p_i.
- For example, the intercept β_0 is the $log(\frac{p_i}{1-p_i})$ for observation i when the predictor takes a value of 0.
- The slope β_1 refers to the change in the expected log odds of failure of an o-ring for a decrease in temperature.

Exercise

Assume that $\log(\frac{p_i}{1-p_i}) = \beta_0 + \beta_1 x_i$.

Show that

$$p_i = \frac{e^{\beta_0+\beta_1 x_i}}{e^{\beta_0+\beta_1 x_i}+1}.$$

This shows that logit function guarantees that the probability p_i lives in $[0, 1]$.

Bayesian Logistic Regression

Recall that

$$Y_i \mid p_i \sim \text{Bernoulli}(p_i) \tag{3}$$

for $i = 1, \ldots, n$.

$$\text{logit}(p_i) := \log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 x_i. \tag{4}$$

How can we build minimal Bayesian prior knowledge?