Stat 343: Warm-up for 20180409

Let's think about two data sets:

Data Set 1: Challenger Space Shuttle O-Rings

On January 28, 1986, the American space shuttle Challenger exploded 73 seconds into flight; all seven crew members on board died. It was later determined that the cause of the explosion was a failure in a joint in one of the booster rockets that launched the shuttle. The shuttle had two booster rockets, each with four sections linked together. The connections between these sections were sealed with O-rings. On the day of the Challenger disaster, one of these O-rings failed to seal the joint, and hot exhaust gas came into contact with unburned fuel, leading to the explosion.

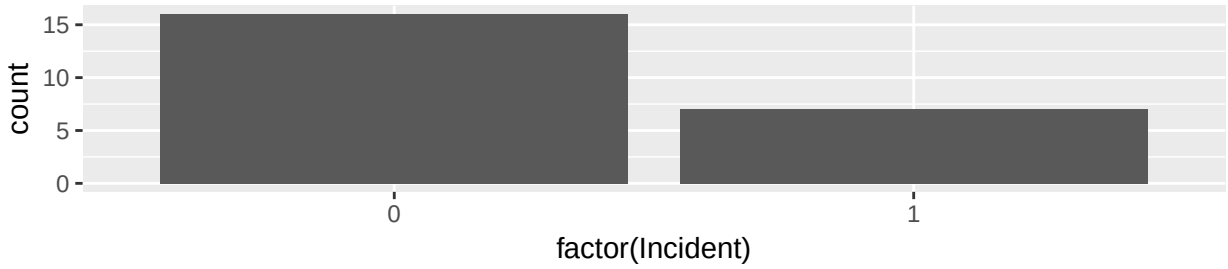
After launch the booster rockets detached from the orbiter and fell into the ocean where they were recovered by NASA, taken apart and analyzed. As part of the analysis NASA recorded whether any of the O-rings were damaged by contact with hot exhaust gas. NASA released the data from all 23 previous launches of the Challenger shuttle with the same booster rocket system. Below are some summaries of the data.

The variable `Incident` is 1 if there was evidence of damage to one of the O-rings on the given shuttle launch, and 0 otherwise.

```
library(tidyverse)
challenger <- read_csv("http://www.evanlray.com/data/chihara_hesterberg/Challenger.csv")
table(challenger$Incident)
```

```
##
##  0  1
## 16  7
```

```
ggplot(data = challenger, mapping = aes(x = factor(Incident))) +
  geom_bar()
```



Define

$$Y_i = \begin{cases} 1 & \text{if there was evidence of damage to on O-ring on launch number } i \\ 0 * & \text{otherwise} \end{cases}$$

What model would you use for $Y_1, \ldots, Y_n$?

What would the maximum likelihood parameter estimate be?

Data Set 2: Loblolly Pines

This example is adapted from Lavine (2002); the description below paraphrases the description of this example in that text.

The amount of carbon dioxide, or CO_2, in the Earth's atmosphere has been steadily increasing over the last century or so. Ecologists are interested in finding out what the impact of this increase in CO_2 might be on Earth's plant life. Carbon is a nutrient needed by plants. It is possible that an increase in CO2 will cause an increase in plant growth which in turn will partly absorb some of the extra carbon in the atmosphere. In order to study this, researchers used a method called Free Air CO_2 Enrichment (FACE). In a FACE experiment, CO2 is released into some plots in a forest. The level of CO2 inside the plot is continually monitored. More CO2 is released as needed to keep the amount of CO2 in the atmosphere at some prespecified level, typically the level that is expected in the mid-21st century.

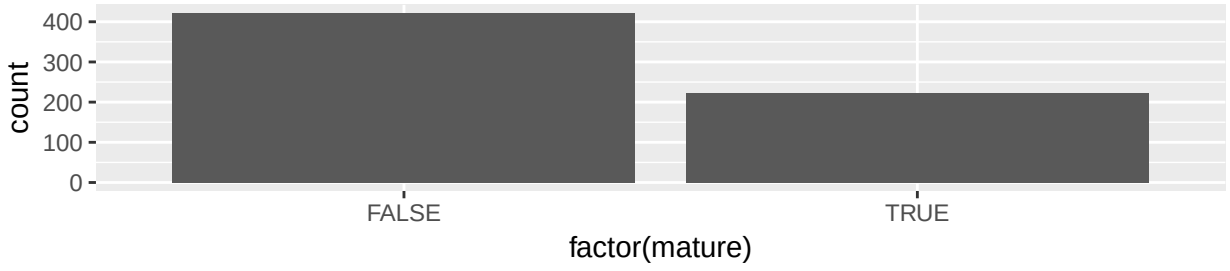
One potental effect of elevated CO_2 is for the trees to reach sexual maturity and hence be able to reproduce earlier than otherwise. Sexually mature trees can produce pine cones but immature trees cannot. So to investigate sexual maturity, a graduate student counted the number of pine cones on each tree.

In the R cell below, we read the data in and create a binary variable called `mature`, with the value 1 if the pine tree produced any pine cones during the three years of the study, and 0 if the pine tree did not produce any pine cones during that time.

```
library(tidyverse)
pinecones <- read_tsv("http://www.evanlray.com/data/lavine_intro_stat_thought/pinecones.txt")
names(pinecones)[7:9] <- c("count_1998", "count_1999", "count_2000")
pinecones <- pinecones %>%
  mutate(
    mature = count_1998 + count_1999 + count_2000 > 0
  )
table(pinecones$mature)
```

```
##
## FALSE  TRUE
##   421   223
```

```
ggplot(data = pinecones, mapping = aes(x = factor(mature))) +
  geom_bar()
```



Define

$$Y_i = \begin{cases} 1 & \text{if pine tree number } i \text{ is mature} \\ 0 * & \text{otherwise} \end{cases}$$

What model would you use for $Y_1, \ldots, Y_n$?

What would the maximum likelihood parameter estimate be?