Fitting and Interpretting Logistic Regression

Logistic regression – Stat 230

In this class, you explore how to model a binary response variable using a logistic regression model. To do this we will analyze the space shuttle O-ring data set described at the beginning of Chapter 7.

You can load the data using the command

```
shuttle <- read.csv("https://aloy.rbind.io/kuiper_data/Shuttle.csv")
```

Describing the data

To begin, let's get an idea of what's in our data set.

Task 1. Complete Chapter 7 Activity 1. "Based on the description of the Challenger disaster O-ring concerns, identify which variable in the `Shuttle` data set in Table 7.1 should be the explanatory variable and which should be the response variable."

Task 2. Complete Chapter 7 Activity 2. "Imagine you were an engineer working for Thiokol Corporation prior to January 1986. Create a few graphs of the [`Shuttle` data]. Is it obvious that temperature is related to the success of the O-rings?"

What's so bad about linear regression?

Before diving into a new model class, let's take a little time to really understand the need for this new model.

Task 3. Complete Chapter 7 Activity 3. "Use the [`Shuttle`] data to create a scatterplot with a least squares regression line for the space shuttle data. Calculate the predicted response values when the temperature is 60°F and when the temperature is 85°F."

Task 4. Create a residual plot for the least squares regression line you just plotted in the previous task. Do the model assumptions appear to be met?

Binary logistic regression

In the previous section you should have seen some pretty obvious deficiencies of the "standard" linear regression model based on a residual plot. In this section, you'll learn how to fit a logistic regression model.

Task 5. To fit a logistic regression model, we use the `glm()` command (which stands for generalized linear model) rather than the `lm()` command. The formula interface is the same, as is the `data` argument, but now we must also add the `family = binomial` argument to specify that we want to fit a logistic regression model. Fill in the blanks below to fit the logistic regression model for the `Shuttle` data set. Compare the estimates for β_0 and β_1 you get from this logistic regression model to those you obtained from simple linear regression.

```
# Fill in the model formula in the first blank
# Fill in the data set name in the 2nd blank
shuttle_glm <- glm(___, data = ___, family = "binomial")
```

Task 6. Complete Chapter 7 Activity 7. "Use Equation (7.9) to predict the probability that a launch has no O-ring damage when the temperature is 31°F, 50°F, and 75°F." You should be able to do this either by hand or with R. To make these predictions in R, we can use the `predict()` function just like in linear regression, but to obtain predicted probabilities (rather than log odds) we have to add the argument `type = "response"`.

```
new_data <- data.frame(Temperature = c(31, 50, 75))
predict(shuttle_glm, newdata = new_data, type = "response")
```

Task 7. Complete Chapter 7 Activity 8. "Calculate the odds of a launch with no O-ring damage when the temperature is 60°F and when the temperature is 70°F." Again you should be able to do this either by hand or with R. To calculate the log odds in R, you can use the `predict()` function but omit the `type = "response"` argument.

Task 8. Complete Chapter 7 Activity 9. "When x_i increases by 10, state in terms of e^{b_1} how much you would expect the log odds to change."

Task 9. Complete Chapter 7 Activity 10. "The difference between the odds of success at 60°F and 59°F is about $0.3285 - 0.2605 = 0.068$. Would you expect the difference between the odds at 52°F and 51°F to also be about 0.068? Explain why or why not.