Inference for Logistic Regression

Logistic regression – Stat 230

In this class, you will conduct inference for single regression coefficients as well as subsets of regression coefficients. To begin, you will continue to 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")
```

Last class you fit a simple logistic regression model where you used the temperature to predict whether a launch would be successful (i.e., there would be no O-ring damage). You used the following code to fit this model:

```
shuttle_glm <- glm(Success ~ Temperature, data = shuttle, family = "binomial")
```

Today, your first tasks are to conduct inference for this fitted model.

Wald intervals and tests

Task 1. Construct a 95% confidence interval for the slope of the logistic regression model. Interpret this interval in context on the odds scale. You can use R to do this, but be sure you are also able to do this "by hand" where you only use R to find z^*, and perhaps as a calculator.

> **Confidence intervals in R**
>
> Wald-based confidence intervals can be constructed in R using the `confint.default()` function. You should pass in your fitted model to this function and can specify the confidence level using the `level` argument (be sure to enter a value between 0 and 1).

Task 2. Complete chapter 7 activity 12. "Calculate the odds ratio of a successful launch between 31°F and 60°F. Provide a confidence interval for this odds ratio and interpret your results." (Notice that this is a 29°F increase in the temperature.) You can use a 95% confidence level for this problem.

Task 3. What happens if we switch what we consider a success and a failure? Activity 13 asks you to explore this idea. To switch what R considers a success you can create a new column, which we'll call `Failure`, that takes on 1 when an O-ring failure occurred and 0 otherwise. The below code

will do this, but if you look at the results you will see a column of `TRUE` and `FALSE` values. This is confusing at first, but R treats `TRUE`s as `1`s and `FALSE`s as `0`s. Use this new column to complete activity 13.

```
library(dplyr)
shuttle <- shuttle |> mutate(Failure = Success == 0)
```

Likelihood ratio tests two ways

In R there are two ways to conduct the likelihood ratio test of $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$. The first is to calculate it "by hand" from the results of the `summary()` function and use the null and residual deviances along with the `pchisq()` function to calculate a p-value (I showed you this approach in the daily prep). The alternative approach is to use the `anova()` command.

> **Likelihood ratio tests in R**
>
> You can carryout likelihood ratio tests in R by fitting a full and reduced model, and then using the `anova()` command to carryout the test.
> For the `shuttle` data, we have already fit the full model. To fit the reduced model with only an intercept we use `1` as our explanatory variable
>
> ```
> shuttle_reduced <- glm( Success ~ 1, data = shuttle, family = "binomial")
> ```
>
> Then, we pass the two models to `anova()` just like in the extra-sums-of-squares F-test, but we add `test = "LRT"` to indicate that this is the likelihood ratio test.
>
> ```
> anova(shuttle_reduced, shuttle_glm, test = "LRT")
> ```

Task 4. Verify that the `anova()` command gives you the same results as the "by hand" calculation of the likelihood ratio test. Once you have verified this, state your conclusion to this test.

Multiple explanatory variables

Now, let's consider a logistic regression model that uses multiple explanatory variables to describe the log odds of success. We'll consider the `Cancer2` data set that is briefly introduced on pages 224-225, and has an expanded introduction in Section 6.1 on page 177. To load the data, run

```
cancer2 <- read.csv("https://aloy.rbind.io/kuiper_data/Cancer2.csv")
```

Task 5. Complete chapter 7 activity 15 (see the book for the prompt). To Fit a multiple logistic regression model in R we add more variables into the formula we pass to `glm()`, just like you added multiple variables to a multiple regression model when you use `lm()`.

Task 6. Complete chapter 7 activity 16.

Task 7. Complete chapter 7 activity 17.

Task 8. Use the drop-in-deviance test (i.e., likelihood ratio test) to compare the two models you fit in tasks 5 and 6 (i.e., activities 15 and 16). State a conclusion to this test in context.

Task 9. Now, conduct a Wald test on the coefficient for `Concave` using the full model. State the test statistic, p-value, and your conclusion. Does your conclusion agree with the drop-in-deviance test? Is the p-value identical?