W271 Assignment 3

Contents


```
library(tidyverse)
library(car)
library(sandwich)
library(lmtest)
library(knitr)
library(Hmisc)
library(gridExtra)
library(stargazer)
library(mcprofile)
```

```
## Warning: package 'mcprofile' was built under R version 4.1.3
```

1 Admission Data: Binary Logistic Regression

(One point per question/sub-question. Eight points total.)

The data set *"admissions.csv"* contains a small sample of graduate school admission data from a university. The variables are specified below:

1. `admit` - the dependent variable that takes two values: $0, 1$ where 1 denotes *admitted* and 0 denotes *not admitted*
2. `gre` - GRE score
3. `gpa` - College GPA
4. `rank` - rank in college major

Suppose you are hired by the University's Admission Committee and are charged to analyze this data to quantify the effect of GRE, GPA, and college rank on admission probability. We will conduct this analysis by answering the following questions:

```
admission <- read_csv('./data/admissions.csv')
```

1.1 Examine the data and conduct an EDA

Examine the data and conduct EDA. Are there any points that are strange, or outlying? Are there any features of the data that affect how you will analyze it?

> 'Fill this in: What do you note about this data?'

1.2 Estimate a logistic regression

Estimate the following binary logistic regressions:

$$
\begin{aligned}
Y &= \beta_0 + \beta_1 GRE + \beta_2 GPA + \beta_3 RANK + e && \text{(Model 1)} \\
Y &= \beta_0 + \beta_1 GRE + \beta_2 GPA + \beta_3 RANK && \text{(Model 2)} \\
&\quad + \beta_4 GRE^2 + \beta_5 GPA^2 + e \\
Y &= \beta_0 + \beta_1 GRE + \beta_2 GPA + \beta_3 RANK && \text{(Model 3)} \\
&\quad + \beta_4 GRE^2 + \beta_5 GPA^2 \\
&\quad + \beta_6 GRE \times GPA + e
\end{aligned}
$$

where $GRE \times GPA$ denotes the interaction between `gre` and `gpa` variables.

```r
model_admission_1 <- 'fill this in'
model_admission_2 <- 'fill this in'
model_admission_3 <- 'fill this in'

## display estimated model in a table

#stargazer(model_admission_1, model_admission_2, model_admission_3, type = "text", omit.stat = "f",
#                    star.cutoffs = c(0.05, 0.01, 0.001), title = "Table 1: The estimated relationship
```

1.3 Test hypotheses

1.3.1 Linear effect: class rank

Using `model_admission_1`, test the hypothesis that class rank has no effect on admission using a likelihood ratio test. Suppose that someone asks, "Are we willing to assume that there is a *linear* effect of class rank as we have in `model_admission_1`?"

> 'Fill this in: What do you observe?'

1.3.2 Linear effect: GRE

Test the hypothesis that $\beta_1 = 0$ in `model_admission_2` using a likelihood ratio test. Interpret what this test result means in the context of a model like what you have estimated in `model_admission_2`.

Then, test the same hypothesis in `model_admission_3` using a likelihood ratio test. Interpret what this test result means in the context of a model like what you have estimated in `model_admission_3`.

1.3.3 Total effect: GRE

Test the hypothesis that GRE has no effect on the likelihood of admission, in a model of admissions defined in `model_admission_3`, using a likelihood ratio test.

1.4 Interpret an effect

Using the entire model, make predictions about how the likelihood of being admitted changes for someone with a $GPA = 3.0$ compared to someone with a $GPA = 4.0$ both with $GRE = 600$.

1.5 Construct a confidence interval

Construct the 95% Profile LR confidence interval for the admission probability for the students with the following profile using **model_admission_3**:

- $GPA = 3.3$;

- $GRE = 720$; and,
- $rank = 1$
- $GPA = 2.5$;
- $GRE = 790$; and,

- $rank = 4$.

```r
admission_probabilty_ci <- 'fill this in'
```

> 'Fill this in: Use inline code evaluation to put all of your written content into the printed page.'

Are the prediction intervals for these two predictions the same? Why or why not? What about the data leads to this similarity or difference?

> 'Fill this in'