W271 Assignment 2

Contents


```
library(tidyverse)

library(sandwich)
library(lmtest)

library(car)
```

1 Placekicking Data: Binary Logistic Regression (3 points – one for each sub-question)

Does the strategy of *icing the kicker* reduce the probability of success for a field goal? The idea is this: In American football, there is a play where a person kicks the ball through the uprights to score points. This is a high-pressure event, and there is a theory that making the kicker stand on the field and think about it will make the kicker nervous, and so make them more likely to miss their attempt.

```
pk <- read_csv('./data/placekick.BW.csv')
```

1.1 Linear Model, with Linear Effects

Use the `distance`, `weather`, `wind15`, `temperature`, `grass`, `pressure` and `ice` as explanatory variables in a logistic regression model that predicts success. Estimate the model, and interpret each of the indicator variables that are used in the model.

1.2 Sun Shine Daydream

The authors use the `Weather==Sun` as the base level category for `Weather`. This is not the default that R uses. Change either the data, or how you estimate the model so that `Weather==Sun` is the base category and other types of weather are the contrasts. Interpret the results.

1.3 Likelihood Ratio Tests

Perform likelihood ratio tests for all explanatory variables to evaluate their importance within the model. Discuss and interpret the results of these tests.

1.4 Should you kick or not?

Suppose that you are trying to make an assessment about whether to kick a field goal in *The Game* – the annual rivalry game played between the Cal Bears and Stanford ... (What is their mascot? A tree?)

Suppose that Cal is down by two points (so `Pressure = Y`), that the distance is 35 yards, and that it is a typical autumn evening in Berkeley, so `Wind15 = 0`, `Weather=Sun`, and `Temperature=Nice`. Cal plays on a turf stadium, and Stanford is out of timeouts, so cannot ice the kicker. What are the chances that Cal makes the kick? Compute the 95% confidence interval.

2 Binary Logistic Regression (4 points – one for each sub-question)

For this question, we use the Mroz data set from *car* library to study factors that are related to married female participation in the labor market.

```
glimpse(Mroz)
```

```
## Rows: 753
## Columns: 8
## $ lfp  <fct> yes, yes, yes, yes, yes, yes, yes, yes, yes, yes, yes, yes, yes, ~
## $ k5   <int> 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ k618 <int> 0, 2, 3, 3, 2, 0, 2, 0, 2, 2, 1, 1, 2, 2, 1, 3, 2, 5, 0, 4, 2, 0,~
## $ age  <int> 32, 30, 35, 34, 31, 54, 37, 54, 48, 39, 33, 42, 30, 43, 43, 35, 4~
## $ wc   <fct> no, no, no, no, yes, no, yes, no, no, no, no, no, no, no, no, no,~
## $ hc   <fct> no, no, no, no, no, no, no, no, no, no, no, yes, yes, no, yes, no~
## $ lwg  <dbl> 1.2101647, 0.3285041, 1.5141279, 0.0921151, 1.5242802, 1.5564855,~
## $ inc  <dbl> 10.910001, 19.500000, 12.039999, 6.800000, 20.100000, 9.859000, 9~
```

In this data set, lfp is a binary variable indicating labor force participation by a married woman during 1975. lfp is equal to one if the woman reports working for a wage outside the home during the year and zero otherwise. We assume that married female labor force participation depends on the following seven potential explanatory variables included in this data set:

- k5: number of kids below the age of 5
- k18: number of kids between 6 and 18
- age: wife's age (in years)
- wc: wife's college attendance
- hc: husband's college attendance
- lwg: log of wife's estimated wage rate
- inc: family income excluding the wife's wage ($1000)

2.1 Estimate a binary logistic regression

Estimate a binary logistic regression with `lfp`, which is a binary variable recoding the participation of the females in the sample, as the dependent variable. The set of explanatory variables includes `age`, `inc`, `wc`, `hc`, `lwg`, `totalKids`, and a quadratic term of `age`, called `age_squared`, where `totalKids` is the total number of children up to age 18 and is equal to the sum of `k5` and `k618`.

```
model_mroz_logit <- 'fill this in'
```

2.2 Evaluate statistical significance

Is the age effect statistically significant?

'Fill this in'

2.3 Interpret an effect

What is the effect of a decrease in age by 5 years on the odds of labor force participation for a female who was 45 years of age.

'Fill this in'

2.4 Construct a confidence interval

Estimate the 95% profile likelihood confidence interval of the probability of labor force participation for females who were 40 years old, had income equal to 20, did not attend college, her husband attend college, had log wage equal to 1, and did not have children.

```
mroz_logit_ci <- 'fill this in'
```

'Fill this in'