Econometrics II TA Session #3

Hiroki Kato

1 Empirical Application of Binary Model: Racial Discrimination in Court

Brief Background. Recently, in the U.S., anti-racism activities called "Black Lives Matter" are getting hot. These activities stems from the death of George Floyd, who was killed by a white police officer on May 25, 2020. The empirical application of binary model investigates whether the judgement of death penalty is based on race of defendant and race of victim.

Data. The package `catdata` contains many built-in dataets which include categorical variables. We use the built-in dataset `deathpenalty` which is about the death-penalty judgement of defendants in cases of multiple murders in Florida between 1976 and 1987.

```
# If there is no package called 'catdata', run 'install.packages("catdata")'
# After that run following codes
library(catdata)
data(deathpenalty)
deathpenalty
```

```
##   DeathPenalty VictimRace DefendantRace Freq
## 1            0          0             0  139
## 2            1          0             0    4
## 3            0          1             0   37
## 4            1          1             0   11
## 5            0          0             1   16
## 6            1          0             1    0
## 7            0          1             1  414
## 8            1          1             1   53
```

Since this datasets is aggregated with repect to `DeathPenalty`, `VictimRace` and `DefendantRace`, we disaggregate it. For example, we make 37 rows whose elements are `DeathPenalty = 0`, `VictimRace = 1`, and `DefendantRace = 0` because there are 37 obaservations, i.e., `Freq = 37`.

```
dt <- deathpenalty
dt <- dt[rep(seq_len(nrow(dt)), dt[,"Freq"]), -4]
```

Model. In a binary model, a dependent (outcome) variable y_i takes only two values, i.e., $y_i \in \{0, 1\}$. A binary variable is sometimes called a *dummy* variable. In this application, the outcome variable is `DeathPenalty` taking 1 if the judgement is death penalty. There are two explanatory variables. First, `VictimRace` is a dummy variable taking 1 if the race of the victim is white. Second, `DefendantRace` is a dummy variable taking 1 if the race of the defendant is white. The regression function is

$$\begin{aligned} &\mathbb{E}[DeathP|Vrace, Drace] \\ =&\mathbb{P}[DeathP = 1|Vrace, Drace] = G(\beta_0 + \beta_1 Vrace + \beta_2 Drace). \end{aligned} \tag{1}$$

The function $G(\cdot)$ is arbitrary function. In practice, we often use following three specifications:

- Linear probability model (LPM): $G(\mathbf{x}_i\beta) = \mathbf{x}_i\beta$.
- Probit model: $G(\mathbf{x}_i\beta) = \Phi(\mathbf{x}_i\beta)$ where $\Phi(\cdot)$ is the standard Gaussian cumulative function.
- Logit model: $G(\mathbf{x}_i\beta) = 1/(1 + \exp(-\mathbf{x}_i\beta))$.