INDE546_HW6

Michael Shieh

2/20/2020

Exercise 1

I want to know how likey will the participants be frustrated with Seattle traffic base on their commute time, how do they feel about the bicyclists in Seattle, and the gender of the participant. Since the dependent variable is categorical but are ordered, I will use **ordered logit regression** to analyze the relationships.

The variable inputed in this model will be:

Dependent variable:

- Frustration (three levels: “Yes”(26.28%), “Neutral”(26.88%), “No”(46.84%))

Independent variables:

- AggresiveBike (three levels: “Agree”(3), “Neutral”(2), “Disagree”(1))
- Commute time (continuous)
- Gender (two levels: “Female”(1), “Male”(2))

Dependent variable:

```
survey <- survey %>%
  rename(Frustration = Please.indicate.how.much.you.agree.or.disagree.with.the.following.statements...I.am.often.frustrated.with.Seattle.traffic.) %>%
  filter(Frustration %in% c("Neither agree nor disagree", "Somewhat disagree", "Somewhat agree", "Strongly agree", "Strongly disagree")) %>%
  mutate(Frustration = case_when(Frustration == 'Neither agree nor disagree' ~ 2, Frustration %in% c("Strongly agree", "Somewhat agree") ~ 3, Frustration %in% c("Strongly disagree", "Somewhat disagree") ~ 1))

survey$Frustration <- factor(survey$Frustration, labels = c("Less", "Neutral", "More"))
survey$Frustration <- relevel(survey$Frustration, ref = "Neutral")

summary(survey$Frustration)
```

```
## Neutral    Less    More
##     135     141     239
```

The above table shows the frequency of all levels in my dependent variable.

Independent variable 1:

I chose this variable because I think if commuters think the bicyclists are too aggressive, they must be bothered by the bicyclists and therefore have the potential to be more frustrated toward the Seattle traffic.

```
survey <- survey %>%
  rename(AggresiveBike = Please.indicate.how.much.you.agree.or.disag
ree.with.the.following.statements...Bicyclists.in.Seattle.are.too.ag
gressive.) %>%
  filter(AggresiveBike %in% c("Neither agree nor disagree", "Somewha
t disagree", "Somewhat agree", "Strongly agree", "Strongly disagree"
)) %>%
  mutate(AggresiveBike = case_when(AggresiveBike == 'Neither agree n
or disagree' ~ 2, AggresiveBike %in% c("Strongly agree", "Somewhat a
gree") ~ 3, AggresiveBike %in% c("Strongly disagree", "Somewhat disa
gree") ~ 1))

survey$AggresiveBike <- factor(survey$AggresiveBike, labels = c("Dis
agree", "Neutral", "Agree"))
survey$AggresiveBike <- relevel(survey$AggresiveBike, ref = "Neutral
")
```

Independent variable 2:

I chose this variable because I think that the longer the commute time, the more opportunity that a commuter would encounter some traffic because they live farther than others.

```
survey <- survey %>%
  rename(CommuteTime = On.average..how.many.minutes.does.it.take.you
.to.get.to.the.U..Washington.from.your.home.) %>%
  filter(CommuteTime >= 0)
```

Independent variable 3:

I chose this variable because I want to see if gender affect the frustration level toward Seattle traffic.

```
survey <- survey %>%
  rename(Gender = Are.you..1) %>%
  filter(Gender %in% c("Male", "Female")) %>%
  mutate(Gender = ifelse(Gender == "Female", 1, 2))
```

Exercise 2

I use polr function to come up with the ordered logit model.

```
ordlog <- polr(Frustration ~ AggresiveBike + CommuteTime + Gender, d
ata = survey, Hess = TRUE)
table <- coef(summary(ordlog))

p <- pnorm(abs(table[, "t value"]), lower.tail = FALSE) * 2
table <- cbind(table, "p value" = p)

rownames(table)[rownames(table) == "AggresiveBikeDisagree"] <- "Disa
gree w/ Aggressive Bicyclist"
rownames(table)[rownames(table) == "AggresiveBikeAgree"] <- "Agree w
/ Aggressive Bicyclist"
rownames(table)[rownames(table) == "CommuteTime"] <- "Commute Time"
#print(xtable::xtable(round(table, 3)), type = "html", html.table.at
tributes = "border=3")

stargazer::stargazer(round(table, 3), type = "html", title = "Summar
y of Ordered Logit Model")
```

Summary of Ordered Logit Model

	Value	Std. Error	t value	p value
Disagree w/ Aggressive Bicyclist	-0.088	0.190	-0.464	0.643
Agree w/ Aggressive Bicyclist	0.496	0.230	2.160	0.031
Commute Time	0.026	0.007	3.568	0
Gender	0.242	0.169	1.427	0.153
Neutral| Less	-0.076	0.324	-0.234	0.815
Less| More	1.116	0.327	3.413	0.001

From the table, I can see that only two variables are significant, and out of the two, commute time is a very significant variable that effects the frustration level toward Seattle traffic.

The following table shows the log likelihood initial and the numbers of parameters.

```
table2 <- cbind("Log Likelihood at Convergence" = deviance(ordlog)/(
-2), "Number of Parameters" = ordlog$edf)
rownames(table2) <- "Value"
print(xtable::xtable(table2), type = "html", html.table.attributes =
"border=3")
```

	Log Likelihood at Convergence	Number of Parameters
Value	-522.25	6.00

Exercise 3

Next, I look at the odds ratios and the corresponding confident intervals of the variable by getting the exponential value of the logit coefficient:

```
round(exp(cbind(OR = coef(ordlog), confint(ordlog))),3)
```

```
## Waiting for profiling to be done...
```

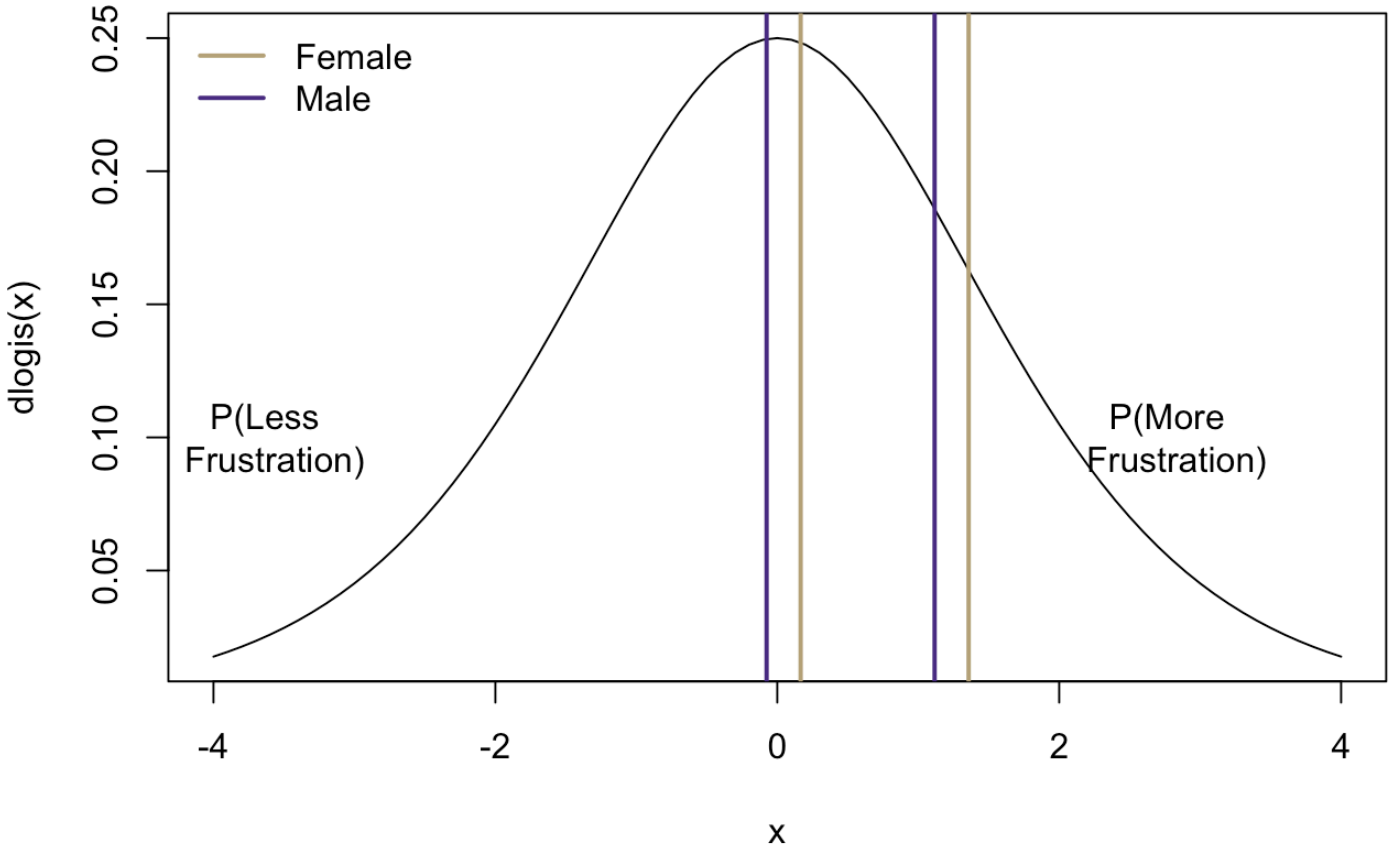
```
##                              OR 2.5 % 97.5 %
## AggresiveBikeDisagree 0.916 0.631  1.328
## AggresiveBikeAgree    1.642 1.050  2.586
## CommuteTime           1.026 1.012  1.041
## Gender                1.273 0.914  1.775
```

As the ratios showed, if you agree that the bicyclists in Seattle are aggressive, it is 1.64 times more likely you will be more frustrated with the Seattle traffic. On the contrary, if you disagree with this aggressiveness, it is 0.92 times less likely you will be less frustrated. Meanwhile, if your commute time increase by 1 unit, it is 1.03 times more likely you will be frustrated with Seattle traffic. Interestingly, it doesn't matter if you hold a positive or negative opinion, as long as you do not have a neutral opinion toward autonomous vehicles, you are more likely to be frustrated with Seattle traffic.

From the following plot, I can say that male have more probability to be frustrated with Seattle traffic compare to female

```
x <- seq(-4, 4, by = .1)

plot(x, dlogis(x), type = "l")
abline(v = c(-0.0757, 1.1162) + 0.2415, col = "#b7a57a", lwd = 2) #F
emale
abline(v = c(-0.0757, 1.1162), col = "#4B2E83", lwd = 2) #Male
text(-3.6, .10, "P(Less \n Frustration)")
text(2.8, .10, "P(More \n Frustration)")
legend("topleft", lwd = 2, legend = c("Female", "Male"), col = c("#b
7a57a", "#4B2E83"), bty = "n")
```



From the following plot, I can say that agreeing to aggressive bicyclists have more probability to be frustrated with Seattle traffic compare to disagreeing to aggressive bicyclists

```
x <- seq(-4, 4, by = .1)

plot(x, dlogis(x), type = "l")
abline(v = c(-0.0757, 1.1162) + 0.4960 - 0.088, col = "#b7a57a", lwd
= 2) #Disagree to Aggression
abline(v = c(-0.0757, 1.1162), col = "#4B2E83", lwd = 2) #Agree to A
ggresion
text(-3.6, .10, "P(Less \n Frustration)")
text(2.8, .10, "P(More \n Frustration)")
legend("topleft", lwd = 2, legend = c("Disagree to Aggression", "Agr
ee to Aggresion"), col = c("#b7a57a", "#4B2E83"), bty = "n")
```