**~~5) Relevant descriptive/exploratory statistics and graphs;~~**

**~~6) Description, assumptions, and justification of statistical methods to be used;~~**

**7) Results of statistical procedures, interpretation/scientific conclusions;**

**~~8) Compare/contrast with findings from other studies; - No Other Studies~~**

**~~9) Discussion of limitations of methods presented, statistical issues, additional scientific questions (and how they might be addressed)~~**

**Project: Using a Proportional Odds Model on Categorical Data**

**PQHS 453**

**Joshua Wu and Wei (David) Dai**

**Summer, 2019**

The purpose of this project is to apply a Proportional Odds Model on Categorical Data. In our case, the categorical variable of interest is a factor, ordered from 8 to 1 (by increments of 1).

**Dataset**

Those with a spinal cord injury have a reduced capability to manage their bladder and bowel movements. Since this is a more than daily activity, it is essential to assess that populations’ opinions on any new technology that might be able to help in those matters. This data comes from a survey and it refers to “nerve stimulation”, a device that gives the person the ability to better manage their bladder and bowels, increasing their quality of life. The data comes from an online survey between 28 June and 29 August 2018. Survey Questions were divided into four categories: demographics, bladder function, bowel function, and attitudes about nerve stimulation.. Our Study population is mainly individual living with spinal cord injury in the US. The variable of interest is an ordinal outcome. It tells us how likely the participant would use an external “Electric Stimulation” device to control their bladder. The predictors will be age, sex, years since of SCI or disorder diagnoses, do you have financial barriers to accessing health care, do you take daily medication to manage your bladder, how much assistance do you usually need to empty your bladder, and do you have a feeling or a body sense of when your bladder needs to be empited.

There is plenty of research on the mechanism of how electric signals allows the user to properly control their bladder. However, there is not much research on the factors that affect whether or not one uses a machine that sends electrical signals. This analysis hopes to answer that gap in knowledge, so that we might be able to allow those who have been diagnosed with a failure in spinal cord activity or have an SCI to have a greater autonomy with the electric signals by finding the relationship between the predictors and the outcome. The hope is to identify the factors promoting or reducing the use of electric signal machines.

The purpose of our analysis is to assess what factors affect whether or not a person with a SCI or a diagnosis will be using a machine that sends electrical signals to control the bladder. Our goal is to use a proportional odds model to identify the factors that promote or reduce the use of electric signal machines.

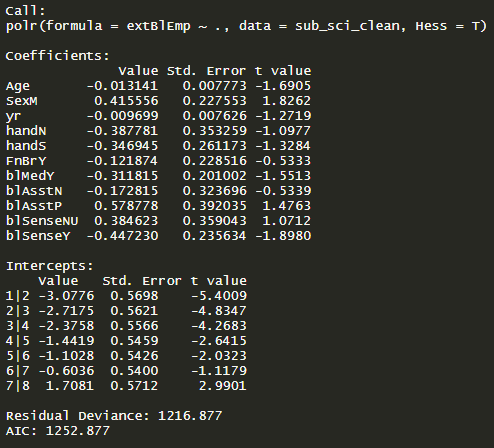
**Proportional Odds Model**

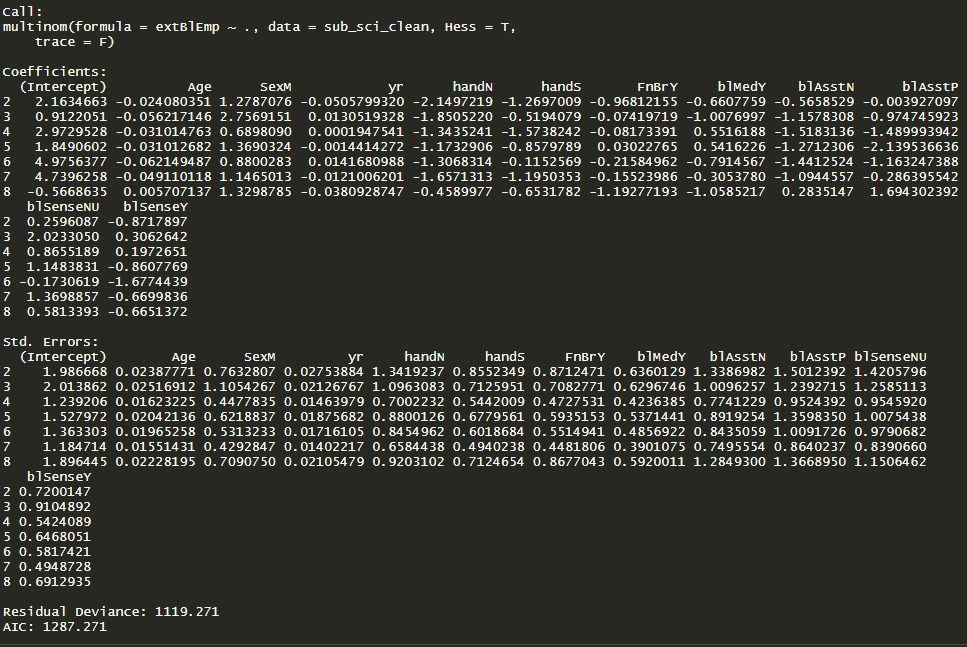
Our categorical outcome is ordinal. That means that there is a clear ordering to the factors of the outcome. In our data, the outcome is split into 8 factors, 1-8. 1-3 equates to “Very Likely” to be using an external ‘electronic stimulus’ device. 4-6 equates to “Somewhat Likely”, while 7-8 equates to “Not Likely”. It is a fluid scale, meaning that the close someone picks a higher number the less likely that they would like to use the external ‘electronic stimulus’ device. The Proportional Odds model requires the proportional odds assumption, which means that for each parameter in the model, the 'slope' estimate between each response level is the same.

**Results**

Proportional Odds or Multinomial:

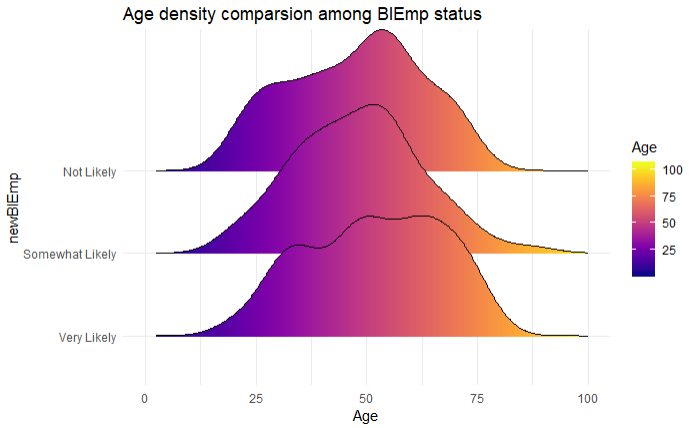
Our data had missing values in it. Therefore we started out by conducting a complete case analysis. We then decided to fit a kitchen sink model for the proportional odds model and a multinomial model to see which model would fit better. We recognize that since our outcome is ordinal that we should go with the proportional odds model regardless of the outcome but we were curious to see. Below are the two fitted models, one is polr, and the other multinom.

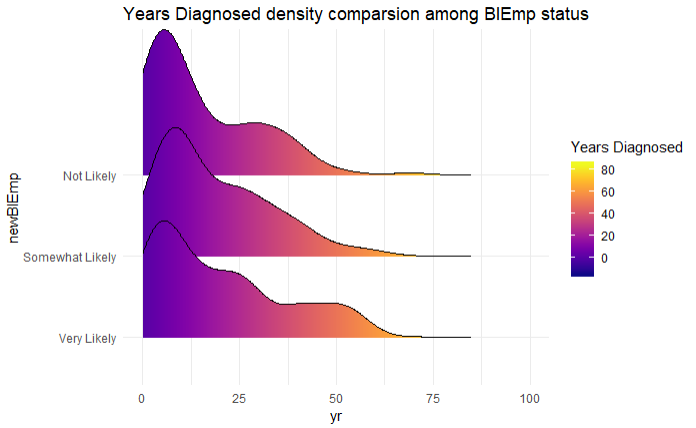


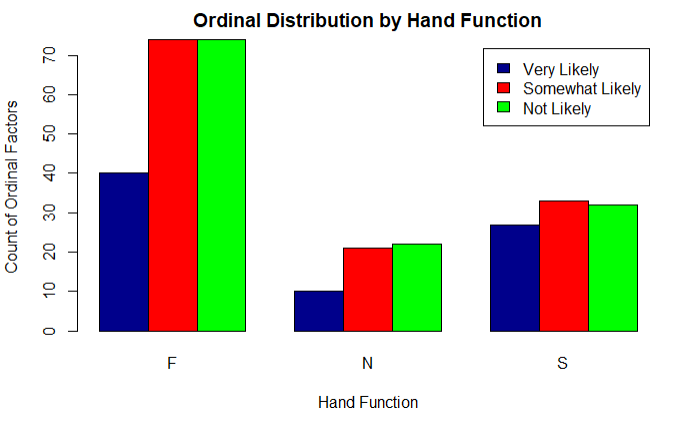
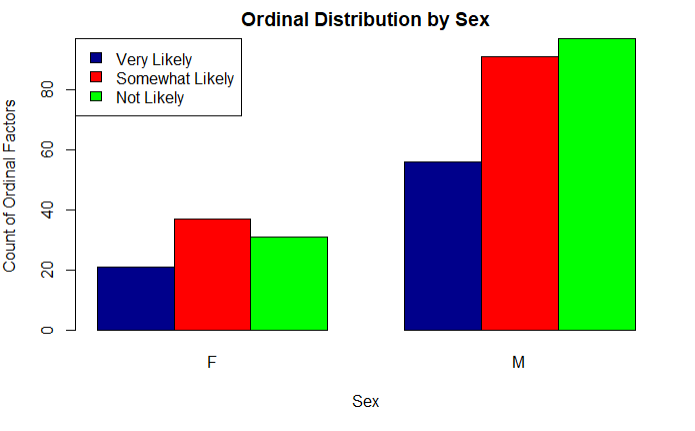
****

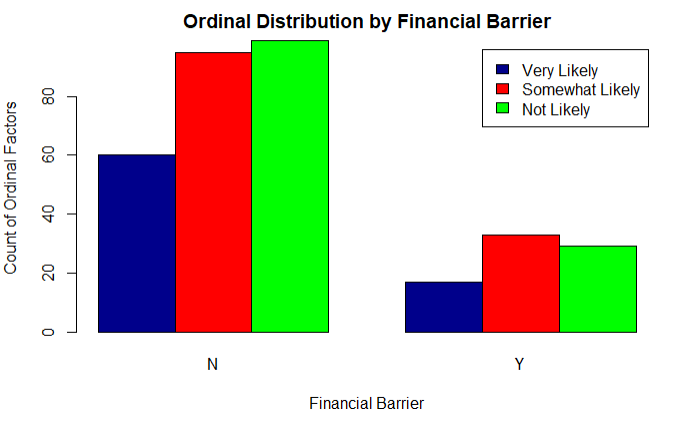
To see if the Proportional Odds Assumption is met, we perform a Likelihood Ratio test between the two. Doing so yields us a p-value of 0.0069, which tells us that the proportional odds model may not fit as well as a standard multinomial mode. Even with this test telling us otherwise, we will still proceeded to use a proportional odds model, since our outcome is an ordinal outcome. We also decided to collapse the 8-level ordinal variable to a 3-level ordinal variable at this point. After collapsing to a 3-level ordinal variable we also found that the proportional odds assumption was valid, with a p-value of 0.21 for the LRT.

Exploratory Data Analysis







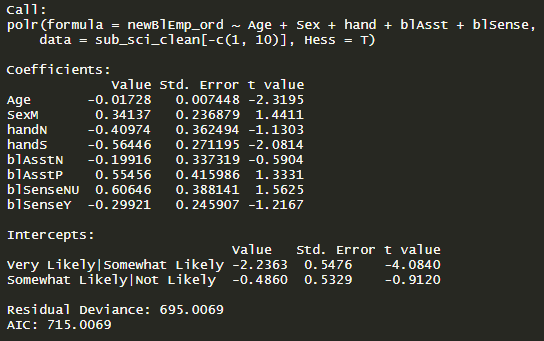


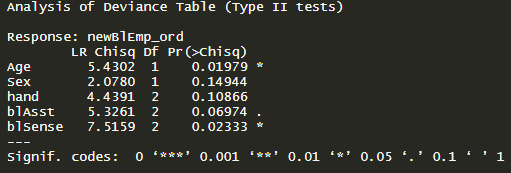
Above are plots of the ordinal factors over each demographic variable: Age, Sex, Years Diagnosed, Hand Function, and Financial Barrier. We felt that it was important to look at the distribution for each of these variables. If we look at Age and Years Diagnosed, we see that the distributions are similar across the three ordinal categories. In addition, it would seem that the distribution of ordinal categories over hand function is similar. We see that there seem to be different distributions for Financial Barrier and Sex.

Variable Selection

For variable selection we chose to perform a stepwise selection on our kitchen sink model. Using AIC as the parameter that would decide whether or not a predictor would stay in the model or not. In our final stepwise selection model we have: Age, Sex, Hand function, Bladder Assistance, and Bladder Sense. The next section contains our final Proportional Odds Model.

Final Result

****



(Interpretations to Follow)

**Comparisons with Other Studies**

We could not find any study that was similar to ours, therefore our analysis will be among the first to analyze various factors and their potential effect on whether or not someone with a reduced capability to control their bladder would be more likely or not to use an external ‘electric stimulus’ device.

**Limitations**

We chose to do a complete case analysis. The strength of that method is that we do not have any missing data, which will not throw off any data. However, by throwing out the participants that had missing data, we lose sample size and those participants. To remedy that, we could have performed a predictive mean matching imputation method to fill in that missing data.

Another limitation of our analysis is that we chose to proceed with a 3-level ordinal model. By condensing levels, we may have lost information

Another limitation of our analysis is the use of stepwise selection to pick our model. While it is quick and simple to obtain a “best fit” model, it also has its limitations. It is possible that the model chosen by stepwise selection is different from the forward selection or the backwards elimination. A way to avoid this could be to avoid automatic selection methods altogether and use another method, such as the LASSO to obtain our “best fit” model.

**Code:**

---

title: "453Proj\_SCIext"

author: "Wei Dai and Joshua Wu"

date: "7/26/2019"

output: html\_document

---

```{r message=FALSE}

library(rms) # multi-logistic regression

library(MASS) # polr

library(nnet) # multinom

library(glmnet) # lasso

library(viridis)

library(ggridges)

library(tidyverse)

```

```{r}

sci\_raw <- read\_csv("curated\_fulldata.csv")

sub\_sci <- sci\_raw %>% select(extBlEmp, Age, Sex, yr, hand, FnBr, blMed, blAsst, blSense)

sub\_sci[c(1, 3, 5:9)] <- lapply(sub\_sci[c(1, 3, 5:9)], as.factor)

summary(sub\_sci)

sub\_sci\_clean <- sub\_sci %>% filter(complete.cases(.)) # remove NAs

```

# preliminary analysis

first step is preliminary analysis: a kitchen sink model of porpotional odds model and multinom model.

## polr

```{r}

mod\_polr <- polr(extBlEmp ~., data = sub\_sci\_clean, Hess = T)

summary(mod\_polr)

```

Some levels could be lumped together 1:3, 4:6, 7:8(intercept are very similar)

## multinom

```{r}

mod\_mtn <- multinom(extBlEmp ~., data = sub\_sci\_clean, Hess = T, trace = F)

summary(mod\_mtn)

```

## LRT

```{r}

pchisq(q = mod\_polr$deviance - mod\_mtn$deviance, df = mod\_mtn$edf - mod\_polr$edf, lower.tail = F)

```

`polr` is not sufficient in describing the data, multinom is required to better fit the data

take away:

\* Some levels of outcome variables could be merged

\* multimon is a better choice than polr

```{r}

# merge into 3 categories

sub\_sci\_clean <- sub\_sci\_clean %>%

mutate(newBlEmp = fct\_recode(extBlEmp, "Very Likely" = "1",

"Very Likely" = "2",

"Very Likely" = "3",

"Somewhat Likely" = "4",

"Somewhat Likely" = "5",

"Somewhat Likely" = "6",

"Not Likely" = "7",

"Not Likely" = "8"),

)

# make it ordered factor

sub\_sci\_clean$newBlEmp\_ord <- factor(sub\_sci\_clean$newBlEmp,

levels = c("Very Likely", "Somewhat Likely", "Not Likely"), ordered = T)

is.ordered(sub\_sci\_clean$newBlEmp\_ord)

```

```{r}

mod\_polr3 <- polr(newBlEmp\_ord ~., data = sub\_sci\_clean[-c(1, 10)], Hess = T)

summary(mod\_polr3)

mod\_mtn3 <- multinom(newBlEmp ~., data = sub\_sci\_clean[-c(1, 11)], Hess = T, trace = F)

summary(mod\_mtn3)

pchisq(q = mod\_polr3$deviance - mod\_mtn3$deviance, df = mod\_mtn3$edf - mod\_polr3$edf, lower.tail = F)

```

# Exploratory data analysis

```{r}

sub\_sci\_clean %>%

ggplot(aes(x = Age, y = newBlEmp, fill = ..x..)) +

geom\_density\_ridges\_gradient() +

scale\_fill\_viridis(option = "C", name = "Age") +

xlim(0, 100) + theme\_minimal() + labs(title = "Age density comparsion among BlEmp status")

```

```{r}

sub\_sci\_clean %>%

ggplot(aes(x = yr, y = newBlEmp, fill = ..x..)) +

geom\_density\_ridges\_gradient() +

scale\_fill\_viridis(option = "C", name = "Years Diagnosed") +

xlim(0, 100) + theme\_minimal() + labs(title = "Years Diagnosed density comparsion among BlEmp status")

```

```{r}

counts <- table(sub\_sci\_clean$newBlEmp\_ord, sub\_sci\_clean$Sex)

barplot(counts, main="Ordinal Distribution by Sex", xlab = "Sex", ylab = "Count of Ordinal Factors", col = c("darkblue", "red", "green"), legend = rownames(counts) ,args.legend = list(x = "topleft"),beside = TRUE)

```

```{r}

counts <- table(sub\_sci\_clean$newBlEmp\_ord, sub\_sci\_clean$hand)

barplot(counts, main="Ordinal Distribution by Hand Function", xlab = "Hand Function", ylab = "Count of Ordinal Factors", col = c("darkblue", "red", "green"), legend = rownames(counts) ,beside = TRUE)

```

```{r}

counts <- table(sub\_sci\_clean$newBlEmp\_ord, sub\_sci\_clean$FnBr)

barplot(counts, main="Ordinal Distribution by Financial Barrier", xlab = "Financial Barrier", ylab = "Count of Ordinal Factors", col = c("darkblue", "red", "green"), legend = rownames(counts) ,beside = TRUE)

```

# Stepwise and Results

```{r}

mod\_polr2 <- polr(newBlEmp\_ord ~., data = sub\_sci\_clean[-c(1, 10)], Hess = T)

mod\_polr2\_step <- step(mod\_polr2, trace = F)

summary(mod\_polr2\_step)

car::Anova(mod\_polr2\_step)

```