POLS6382 Quantitative Method III
Maximum Likelihood Estimation
Lab 5: Models for Ordinal Data

Ling Zhu and Emily Lee
Department of Political Science
University of Houston

2025/10/08

1. Learning Objectives

- Learn how to estimate a model suitable for ordinal outcomes.

- Learn how to substantively interprets the statistical results of an ordered logit model.

```
> rm(list=ls())
> setwd("/Users/lingzhu/Dropbox/UH Teaching/POLS6382_2025/2025 Labs/Lab 5")
> my_packages <- c("foreign", "ggplot2", "gmodels", "Hmisc","MASS", "ordinal", "reshape2","stats")
> invisible(lapply(my_packages, require, character.only = TRUE))
> healthcare<-read.dta("gss2012lab5.dta")
> attach(healthcare)
```
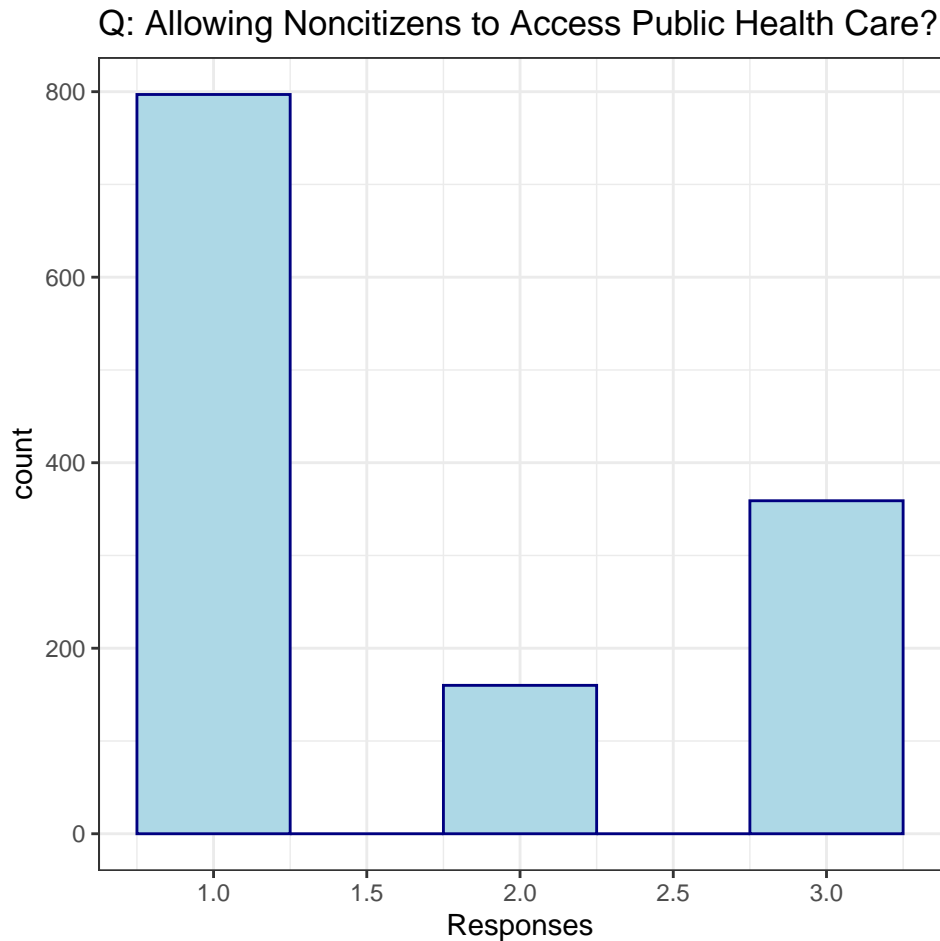
2. Data Example: Americans' Preferences on Immigrants' Access to Public Health Care

In this section, we will learn how to estimate an ordered logistic regression model, using the example data, `gss2012lab5.dta`. Data are drawn from the 2012 General Social Survey (GSS). In the datafile, we have the following variables. For more details about the dataset, see Zhu, Ling and Kenicia Wright. 2016. "Why Do Americans Dislike Publicly Funded Health Care? Examining the Intersection of Race and Gender in the Ideological Context." *Politics, Groups, and Identities*, 4(4): 618-637.

- `hlthctzen`: Respondents' answers to the question: *Do you agree with the following statement?One should have access to public funded health care if he/she is not a citizen.* This variable is coded as three categories: "1"= disagree, "2"=neither, and "3"= agree with the statement.

- `female`: respondents' gender, coded as "1" for female respondents, and "0" otherwise.

- `white`: Respondents' race, coded as "1" for whites and "0" otherwise.

- `age`: Respondents' age.

- `educ`: Year of education.

- `income`: measured as levels: 1=1-$1,000, 2=$1,00-2,999, 3 =$3,000-3,999....12=$25,000 or more.

- `partyid`: party identification, measured as levels: 0= strong Democrat, 1=not strong Democrat, 2= independent, near democrat, 3=independent, 4=independent, near Republican, 5=not strong Republican, 6=strong Republican.

- `ideology`: liberal-conservative ideology scale, coded as a 1-to-7 scale, reflecting conservatism.

The variable of interest is `hlthctzen`, respondents' opinions about whether non-citizens should have access to public health care or not. We may just use some descriptive statistics to learn about the sample distribution. A histogram plot could be useful to describe how preferences vary. From Figure 1 we see that the total sample size is about 1,300 cases. About 800 respondents disagree with the statement that a non-citizen should have access to public health care. About 400 respondents agreed with the statement.

```
> ggplot(healthcare, aes(x=hlthctzen)) +
+   geom_histogram(binwidth=0.5, fill="lightblue",color="navy")+
+   labs(title="Q: Allowing Noncitizens to Access Public Health Care?", x="Responses")+
+   theme_bw()
```


We can also obtain marginal proportions for each choice category using R function, `prob.  table()`. Based on the 2012 GSS, 60.6% of the respondents disagreed with the statement. About 12% of the respondents reported neutral attitudes, and only 27.3% of the respondents agreed that non-citizens should have access to public health care. In other words, the majority opinion is that non-citizens should be excluded from accessing public health care.
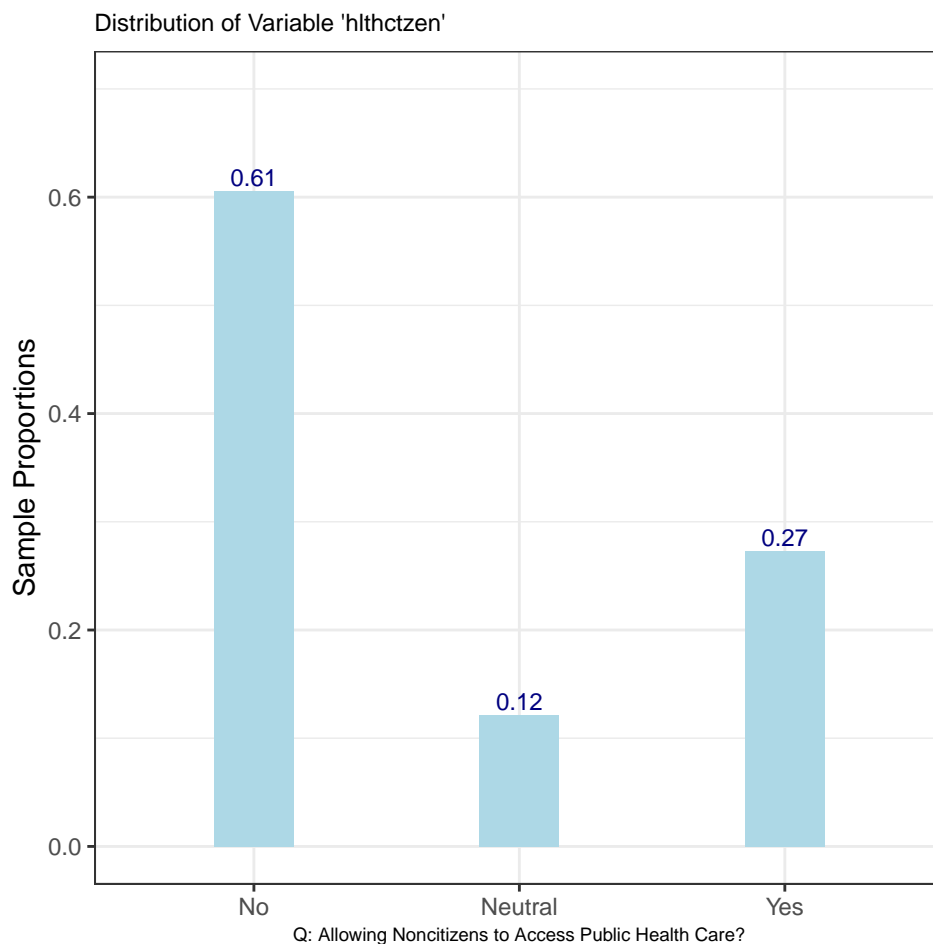
```
> data.frame(table(hlthctzen))

  hlthctzen Freq
1         1  797
2         2  160
3         3  359
> prop.table(table(hlthctzen))
```

```
hlthctzen
        1         2         3
0.6056231 0.1215805 0.2727964
```

To visualize sample proportions, we can convert the sample frequencies into proportions first, then plot the sample proportions in a bar chart.

```
> data2<-as.data.frame(prop.table(table(hlthctzen)))
> ggplot(data2, aes(y=Freq, x=hlthctzen)) +
+ geom_bar(stat="identity", width=0.3, fill="lightblue")+
+ ggtitle("Distribution of Variable 'hlthctzen'") +
+ labs(y="Sample Proportions",
+ x="Q: Allowing Noncitizens to Access Public Health Care?") +
+ theme_bw() +ylim(0,0.7)+
+ geom_text(aes(label = round(Freq,2)), vjust = -0.3, size=3, color="navy")+
+ scale_x_discrete(breaks = c("1","2","3"), labels=c("No","Neutral","Yes")) +
+ theme(axis.title.x = element_text(size = 7),
+ plot.title = element_text(size = 9))
```


3. Estimating an Ordered Logit Model with `polr()`

Various `R` functions can be used to estimate an ordered logit/probit model. One commonly used function is `polr()` from package `MASS`. Estimation with `polr()` is simple. We define a model equation, and the data file. The default link function is "logit". The statement `Hess=TRUE` returns the Hessian matrix in model object,

which will be used in post-estimation simulation later. We specify our ordered-logit model by including `gender`, `race`, `age`, `education`, `income`, and `ideology` as the explanatory variables.

```
> model1<-polr(as.factor(hlthctzen)~female+white+age+educ+income+ideology,
+                  data=healthcare,Hess=TRUE)
> summary(model1)

Call:
polr(formula = as.factor(hlthctzen) ~ female + white + age +
    educ + income + ideology, data = healthcare, Hess = TRUE)

Coefficients:
              Value Std. Error t value
female     0.0786593   0.118059  0.6663
white     -1.0827232   0.135465 -7.9926
age       -0.0009609   0.003554 -0.2704
educ       0.0358809   0.020647  1.7378
income    -0.0156195   0.029821 -0.5238
ideology  -0.3705436   0.042046 -8.8129


Intercepts:
     Value    Std. Error t value
1|2 -1.5645   0.4641     -3.3707
2|3 -0.9415   0.4631     -2.0332


Residual Deviance: 2232.765
AIC: 2248.765
```

You may notice from the above R output that `polr()` does not return p-values. We can look at the t-value column to evaluate if a variable has a significant effect on respondents' preferences regarding non-citizen's access to health care. You can also get p-values using `pnorm()`.

```
> #store table
> (ctable<-coef(summary(model1)))

                  Value Std. Error     t value
female     0.0786593115 0.11805945   0.6662687
white     -1.0827231790 0.13546543  -7.9926159
age       -0.0009608995 0.00355360  -0.2704017
educ       0.0358808677 0.02064692   1.7378316
income    -0.0156194898 0.02982057  -0.5237824
ideology  -0.3705435895 0.04204551  -8.8129176
1|2       -1.5644559949 0.46414069  -3.3706504
2|3       -0.9415444911 0.46307425  -2.0332474

> # get p values
> p<-pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2
> #combined table
> (ctable <- cbind(ctable, "p value" = p))

                  Value Std. Error     t value      p value
female     0.0786593115 0.11805945   0.6662687 5.052394e-01
white     -1.0827231790 0.13546543  -7.9926159 1.321053e-15
age       -0.0009608995 0.00355360  -0.2704017 7.868512e-01
educ       0.0358808677 0.02064692   1.7378316 8.224049e-02
income    -0.0156194898 0.02982057  -0.5237824 6.004299e-01
ideology  -0.3705435895 0.04204551  -8.8129176 1.219302e-18
```

```
1|2        -1.5644559949 0.46414069 -3.3706504 7.499097e-04
2|3        -0.9415444911 0.46307425 -2.0332474 4.202754e-02
```

4. Substantive Interpretation

How can we substantively interpret the coefficients reported in Table 1? Scott Long discussed various methods in his textbook. Here, we will focus on using odds ratios and predicted probabilities.

4.1 Odds Ratio

Because we estimated an ordered logit model, we can convert coefficients into odds ratios, using function `exp()`. Combining `exp()` with `confint()`, we can also obtain confidence intervals for odds ratios. If the confidence intervals of an odds ratio go across 1, it means that the odds ratio coefficient is not statistically significant.

```
> exp(coef(model1))

   female     white       age      educ    income  ideology
1.0818357 0.3386720 0.9990396 1.0365324 0.9845019 0.6903590
```

We observe that `white` and `ideology` significantly affect respondents' preferences regarding non-citizens' access to public health care.

1. The odds ratio associated with `white` is 0.34: the odds of reporting "neither" and "agree" versus "disagree" are 66 percent lower for white respondents than that for non-white respondents. In other words, compared with non-white respondents, white respondents are more likely to choose "disagree" than "neither" and "agree".

2. `ideology` has an odds ratio of 0.69: a one-point increase in conservatism decreases the probability of choosing "agree" over "neither" and "disagree" by about 31%. In other words, moving from liberalism to conservatism will decrease the probability of choosing "agree" over the other two lower categories.
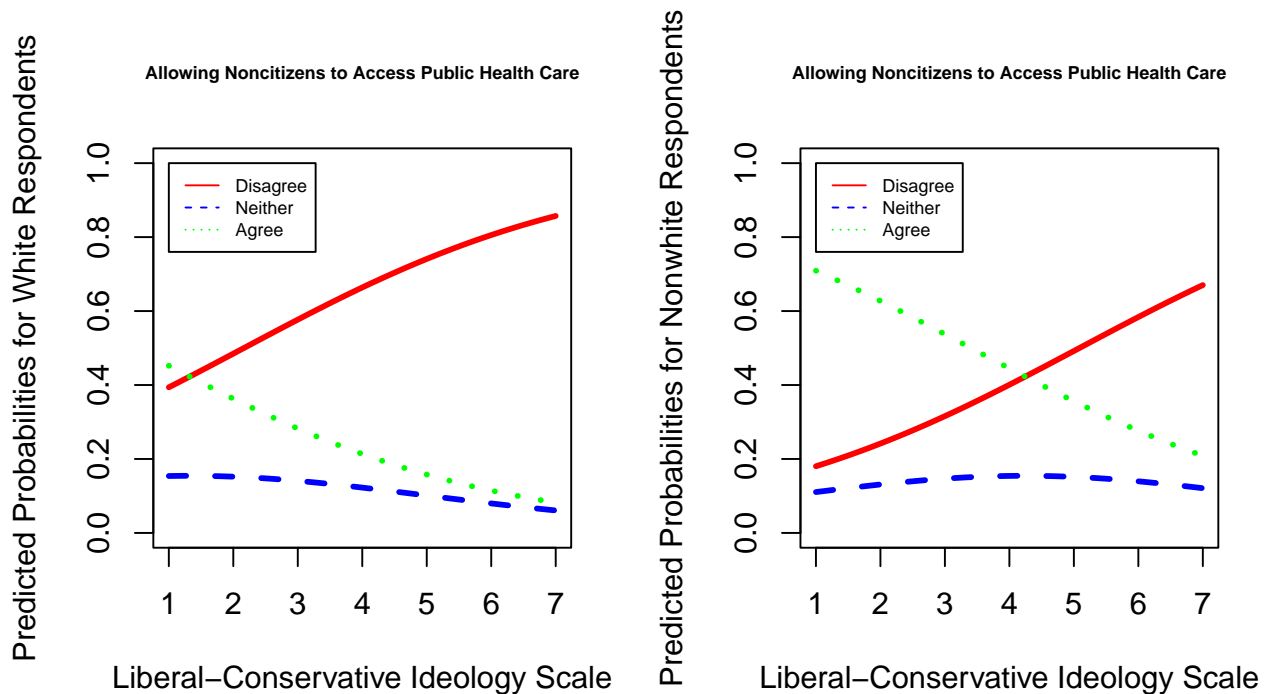
4.2 Predicted Probabilities

Beyond using odds ratios, another useful way to substantively interpret findings is to calculate predicted probabilities associated with each choice category. First, we will discuss how to present mean predictions. Suppose that we are interested in gauging the substantive effect of `ideology` and `race` (being white) on preferences. What we can do is to predict the probabilities of choosing each choice category across the full range of the ideology scale, but setting the dummy variable `white` to be 1 and 0. This would give us two sets of predictions: one for white respondents, and one for non-white respondents. Note that we hold all the other variables constant. `female=mean(female)` essentially gives variable `female` an artificial value. This is to average the weights between female and male respondents in the sample. Alternatively, you can set `female` to be 0 or 1. That will give you the comparison between white male and non-white male respondents (or white and non-white female respondents) across the full range of `ideology`.

```
> # Mean predicted probabilities, white respondents
> predictdata<-cbind(ideology=seq(1,7,length=100),
+                    age=mean(age),income=mean(income),educ=mean(educ),
+                    female=mean(female),white=1)
> opinion.hat<-predict(model1,predictdata,type='prob')
>
> # Mean predicted probabilities, non-white respondents
> predictdata2<-cbind(ideology=seq(1,7,length=100),
+                    age=mean(age),income=mean(income),educ=mean(educ),
+                    female=mean(female),white=0)
> opinion.hat2<-predict(model1,predictdata2,type='prob')
>
```

```
> # Plot pps
> par(mfrow=c(1,2))
> ideology<-seq(1,7,length=100)
> plot(c(1,7),c(0,1),type='n',
+     xlab="Liberal-Conservative Ideology Scale",
+     ylab="Predicted Probabilities for White Respondents",cex=0.8,
+     main="Allowing Noncitizens to Access Public Health Care", cex.main=0.6)
> lines(ideology,opinion.hat[1:100,1],lty=1,lwd=3,col="red")
> lines(ideology,opinion.hat[1:100,2],lty=2,lwd=3,col="blue")
> lines(ideology,opinion.hat[1:100,3],lty=3,lwd=3,col="green")
> legend(1,1,cex=0.6,c('Disagree','Neither','Agree'),
+        lty=1:3,col=c("red","blue","green"))
> plot(c(1,7),c(0,1),type='n',
+     xlab="Liberal-Conservative Ideology Scale",
+     ylab="Predicted Probabilities for Nonwhite Respondents",cex=0.7,
+     main="Allowing Noncitizens to Access Public Health Care", cex.main=0.6)
> lines(ideology,opinion.hat2[1:100,1],lty=1,lwd=3,col="red")
> lines(ideology,opinion.hat2[1:100,2],lty=2,lwd=3,col="blue")
> lines(ideology,opinion.hat2[1:100,3],lty=3,lwd=3,col="green")
> legend(1,1,cex=0.6,c('Disagree','Neither','Agree'),
+         lty=1:3,col=c("red","blue","green"))
```
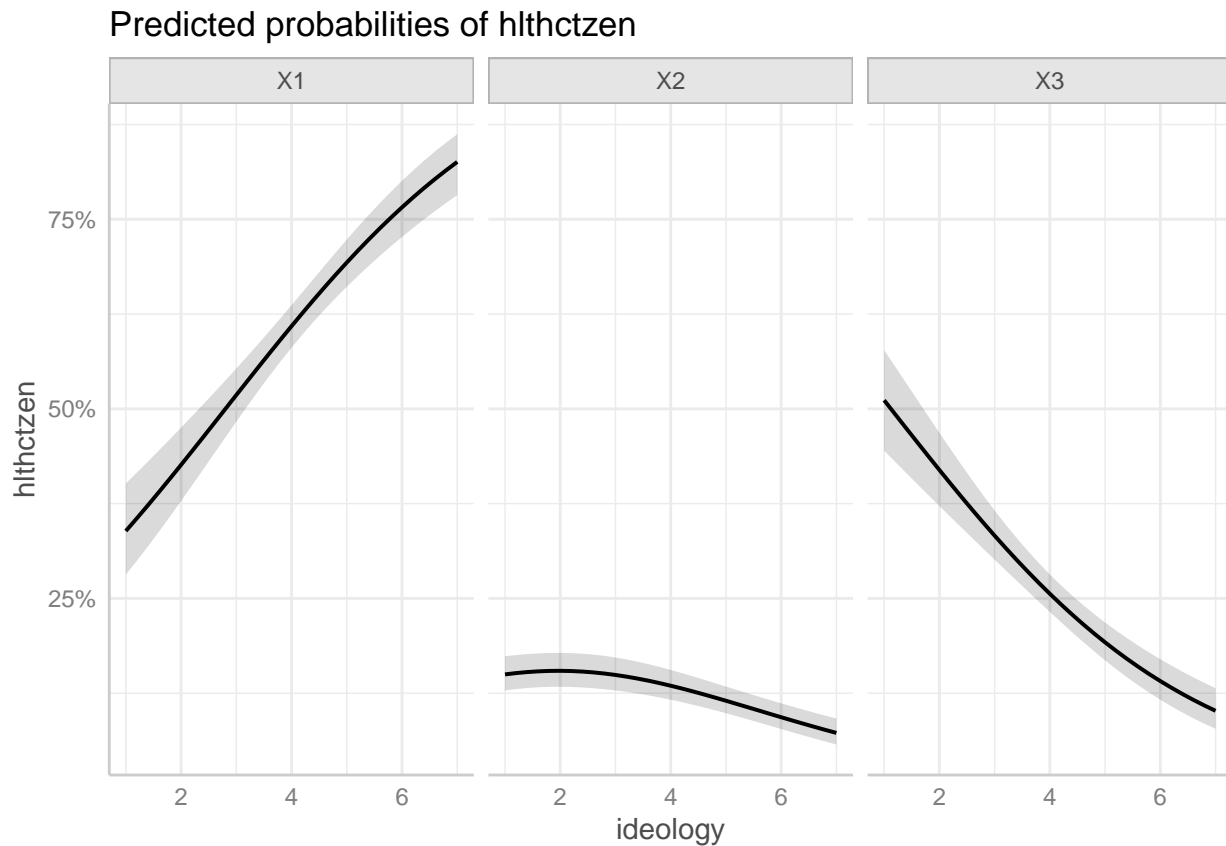

We observe that `ideology` has a negative impact on supporting non-citizens' access to public health care, which is consistent across the two racial groups. For white respondents, conservatism increases the predicted probability of choosing "disagree" with non-citizens having access to public health care, while substantially decreases the predicted probability of choosing "agree" with non-citizens having access to public health care. A similar pattern is observed for non-white respondents that conservatism increases the probability of opposing noncitizens' access to public health care.

```
> # Appendix: Using ggeffects package
> require(tidyverse)
> require(ggeffects)
```

```
> ggeffect(model1, terms = "ideology[1:7 by = 0.2]") %>%
+    plot()
```

Predicted probabilities of hlthctzen


```
> prediction_data<-ggeffect(model1, terms = "ideology[1:7 by = 0.2]")
> ggplot(data = prediction_data, aes(x = x, y = predicted,
+                 color = response.level, group = response.level))+
+    geom_line() +
+    #geom_point(alpha=0.8) +
+    geom_ribbon(aes(ymin = conf.low, ymax = conf.high,
+                    color = response.level,
+                    group = response.level),
+                    alpha=0.2) +
+    scale_color_brewer(palette = "Dark2",
+                    name = "",
+                    labels = c("Oppose",
+                                "Neutral",
+                                "Support")) +
+    labs(
+      x = "Ideology (Conservatism))",
+      y = "Attitudes toward Non-Citizen Access to Public Health Care") +
+    theme_bw() +
+    theme(
+      legend.position = "bottom", # move legend to the bottom
+      axis.title = element_text(size = 14) # increase axis title size
+        )
```