Multinomial Logistic Regression

Note: The dataset used in this tutorial and the R Script are on Moodle):

Loading the 2016 CCES dataset

```
install.packages("foreign", dependencies=TRUE)
library(foreign)

dat <- read.dta(file.choose(), convert.factors=FALSE)
```

Recode Variables Quickly

1. Here, we recode a few variables that we will need later on.

```
table(dat$CC16_340a)
dat$ideology <- recode(dat$CC16_340a, "8=NA")
summary(dat$ideology)

table(dat$religpew)
dat$religion <- recode(dat$religpew, "1:2='Protestant-Catholic'; 3='Mormon';
4='Orthodox'; 5='Jewish'; 6:8='Other';  9='Atheist'; 10:11='Nothing'; else=NA")
table(dat$religion)

dat$religion <- factor(dat$religion, levels=c("Protestant-Catholic", "Mormon", "Orthodox",
"Jewish", "Other", "Nothing", "Atheist"))

table(dat$birthyr)
dat$age <- 2016 - dat$birthyr
table(dat$age)
summary(dat$age)

table(dat$faminc)
dat$income <- recode(dat$faminc, "31=NA; 97=NA")
table(dat$income)
summary(dat$income)

table(dat$gender)
dat$gender1 <- recode(dat$gender, "1=0; 2=1")
table(dat$gender)

table(dat$race)
dat$race1 <- recode(dat$race, "1='White'; 2='Black'; 3='Hispanic'; 4:8='Other'")
table(dat$race1)
```

```
dat$race1 <- factor(dat$race1, levels=c("White", "Black", "Hispanic", "Other"))

table(dat$CC16_410a)
dat$votechoice <- recode(dat$CC16_410a, "1='1'; 2='0'; else=NA")
table(dat$votechoice)
```

Multivariate Multinomial Logistic Regression

1. Finally, what if we have a dependent variable that was nominal level. Here, we would utilize multinomial logistic regression. When conducting multinomial logistic regression we set one category as the reference category, and we compare whether coefficients for the other categories are statistically different. For example, let us say we wanted to predict vote choice in the last election, but we want to explore more than just voting for Trump or Clinton.

```
table(dat$CC16_410a)

dat$votechoice2 <- recode(dat$CC16_410a, "1='Trump'; 2='Clinton'; 3='Johnson';
4='Stein'; else=NA")

multi.mod <- multinom(votechoice2 ~ age + gender1 + income + religion + race1 +
ideology, data=dat)

summary(multi.mod)

smulti <- summary(multi.mod)
multi.t <- smulti$coefficients/smulti$standard.errors
multi.p <- pt(abs(multi.t),
  nrow(multi.mod$fitted.values)-multi.mod$edf,
  lower.tail=F)
b <- round(smulti$coefficients, 3)
b[which(multi.p > .05, arr.ind=T)] <- ""

noquote(b)
```

Figure 1: Multinomial Model Predicting Importance of Immigration

```
> summary(multi.mod)
Call:
multinom(formula = votechoice2 ~ age + gender1 + income + religion +
    race1 + ideology, data = dat)

Coefficients:
        (Intercept)          age    gender1       income religionMormon
Johnson   -2.953211 -0.034520806 -0.6108011 -0.003374058     0.72438672
Stein     -1.781903 -0.018898280 -0.2074477 -0.106079852     0.02563869
Trump     -4.712224  0.003423348 -0.3053432 -0.011732117     0.05472020
        religionOrthodox religionJewish religionOther religionNothing
Johnson       -0.3881167     -0.3621400    -0.6300632       0.1128498
Stein          0.4204916     -0.6100021     0.4935252       0.4933716
Trump         -0.0404126     -0.8488740    -1.0299102      -0.2877280
        religionAtheist race1Black race1Hispanic race1Other   ideology
Johnson     -0.03941726 -2.4201727    -0.8961068 -0.2301748 0.73737532
Stein        0.62162023 -0.8903184    -0.1013619  0.2380760 0.01082681
Trump       -1.06880661 -3.4377386    -1.1457819 -0.3957655 1.20216361

Std. Errors:
        (Intercept)         age    gender1      income religionMormon
Johnson  0.14952228 0.001889648 0.05734276 0.008769092      0.1982240
Stein    0.19054388 0.002536854 0.07902866 0.012461350      0.5141588
Trump    0.09592744 0.001057665 0.03238301 0.004993082      0.1444462
        religionOrthodox religionJewish religionOther religionNothing
Johnson        0.4058665     0.17876747     0.2452785      0.06580388
Stein          0.5183008     0.32639472     0.2253110      0.09115508
Trump          0.1932078     0.09649962     0.1486719      0.03849145
        religionAtheist race1Black race1Hispanic race1Other   ideology
Johnson      0.10832169 0.15783108    0.11268161 0.10534845 0.02012560
Stein        0.11804723 0.14879357    0.13823641 0.12959071 0.02874256
Trump        0.08438318 0.08501571    0.06300361 0.06405772 0.01337088

Residual Deviance: 42548.97
AIC: 42632.97
>
> smulti <- summary(multi.mod)
> multi.t <- smulti$coefficients/smulti$standard.errors
> multi.p <- pt(abs(multi.t),
+   nrow(multi.mod$fitted.values)-multi.mod$edf,
+   lower.tail=F)
> b <- round(smulti$coefficients, 3)
> b[which(multi.p > .05, arr.ind=T)] <- ""
>
> noquote(b)
        (Intercept) age    gender1 income religionMormon religionOrthodox
Johnson -2.953      -0.035 -0.611         0.724
Stein   -1.782      -0.019 -0.207  -0.106
Trump   -4.712      0.003  -0.305  -0.012
        religionJewish religionOther religionNothing religionAtheist race1Black
Johnson -0.362         -0.63         0.113                           -2.42
Stein   -0.61          0.494         0.493           0.622           -0.89
Trump   -0.849         -1.03         -0.288          -1.069          -3.438
        race1Hispanic race1Other ideology
Johnson -0.896        -0.23      0.737
Stein                 0.238
Trump   -1.146        -0.396     1.202
>
```

There is a new easier way to summarize these models using the stargazer package that we have been using.

```
stargazer(multi.mod)
```

Table 1:

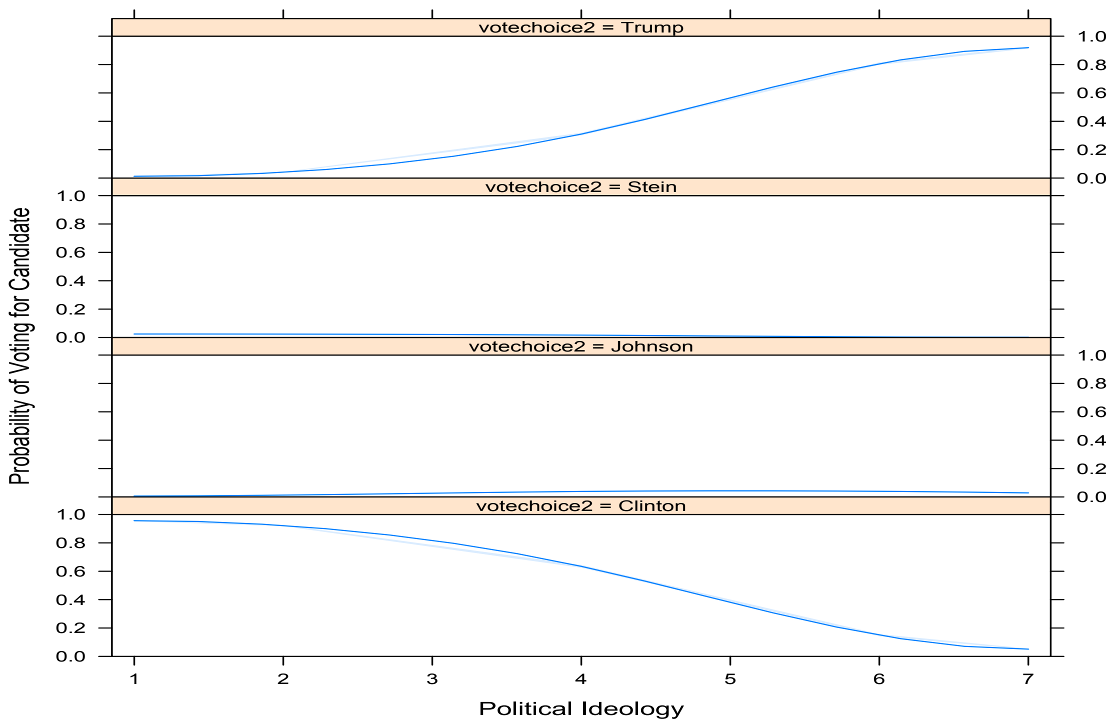
	Dependent variable:		
	Johnson	Stein	Trump
Constant	-2.953^{***}	-1.782^{***}	-4.712^{***}
	(0.150)	(0.191)	(0.096)
Age	-0.035^{***}	-0.019^{***}	0.003^{***}
	(0.002)	(0.003)	(0.001)
Woman	-0.611^{***}	-0.207^{***}	-0.305^{***}
	(0.057)	(0.079)	(0.032)
Income	-0.003	-0.106^{***}	-0.012^{**}
	(0.009)	(0.012)	(0.005)
Mormon	0.724^{***}	0.026	0.055
	(0.198)	(0.514)	(0.144)
Orthodox	-0.388	0.420	-0.040
	(0.406)	(0.518)	(0.193)
Jewish	-0.362^{**}	-0.610^{*}	-0.849^{***}
	(0.179)	(0.326)	(0.096)
Other	-0.630^{**}	0.494^{**}	-1.030^{***}
	(0.245)	(0.225)	(0.149)
Nothing	0.113^{*}	0.493^{***}	-0.288^{***}
	(0.066)	(0.091)	(0.038)
Atheist	-0.039	0.622^{***}	-1.069^{***}
	(0.108)	(0.118)	(0.084)
Black	-2.420^{***}	-0.890^{***}	-3.438^{***}
	(0.158)	(0.149)	(0.085)
Hispanic	-0.896^{***}	-0.101	-1.146^{***}
	(0.113)	(0.138)	(0.063)
Other	-0.230^{**}	0.238^{*}	-0.396^{***}
	(0.105)	(0.130)	(0.064)
Political Ideology	0.737^{***}	0.011	1.202^{***}
	(0.020)	(0.029)	(0.013)
N			36,041
Akaike Inf. Crit.			42,632.970
Note:			*p<0.1; **p<0.05; ***p<0.01

We can even plot the effects like we did before in the analysis with the bivariate dependent variable.

```
eff1 <- effect("ideology", multi.mod, default.levels=100, typical=median)

effect1 <- print(plot(eff1, rescale.axis=F, rug=FALSE, xlab="Political Ideology",
ylab="Probability of Voting for Candidate", main="", ylim=c(0,1)))
```

Figure 2: Effect of Ideology on Vote Choice



Lab Activity

In the 2020 Finland European Social Survey dataset, find a nominal level variable that you are interested in. Recode the variable and estimate a model with socio-demographic variables as the predictors. Examples: religious denomination, marital status, etc.