Nominal Dependent Variables

Week 19

Winnie Xia

Nominal Data in Political Science

- If outcome or dependent variable is binary, then use **logit or probit models**.
- If outcome or dependent variable is categorical but ordered (i.e low to high), then use **ordered logit or ordered probit models**
  - Agree; Neutral; Disagree
  - Low; Middle; High
- If outcome or dependent variable is categorical without any particular order, then use **multinomial logit/probit models**.

Models for Nominal DVs

- Multinominal logit model: Series of binary logit models; similar to ordered logit. **Easiest to interpret**.
- Multinominal probit model: Series of binary probit models; similar to ordered probit. **Drops the IIA assumption**.
- Conditional logit model
- Joint multinominal and conditional logit model
- Mixed multinomial logit model
- Nested multinomial logit model

R packages for Nominal families

	nnet	mlogit	mnlogit	MNP	mclogit
Multinominal logit model	yes	yes	yes	no	yes
Multinominal probit model	no	yes	no	yes	no
Conditional logit model	no	yes	yes	no	yes
Joint multinominal and conditional logit model	no	yes	no	no	yes
Mixed multinomial logit model	no	yes	yes	no	no
Mixed multinomial logit model	no	no	yes	no	no

Multinomial Logit Model in R

```r
library(foreign)
library(texreg)
dat <- read.dta("https://stats.idre.ucla.edu/stat/data/hsbdemo.dta")
head(dat)
```

```
##    id female    ses schtyp     prog read write math science socst       hon
## 1  45 female    low public vocation   34    35   41      29    26 not enrol
## 2 108   male middle public  general   34    33   41      36    36 not enrol
## 3  15   male   high public vocation   39    39   44      26    42 not enrol
## 4  67   male    low public vocation   37    37   42      33    32 not enrol
## 5 153   male middle public vocation   39    31   40      39    51 not enrol
## 6  51 female   high public  general   42    36   42      31    39 not enrol
```

Multinomial Logit Model in R

```r
library(nnet)
dat$prog2 <- relevel(dat$prog, ref = "academic")
m1 <- multinom(prog2 ~ ses + write, data = dat)
```

```
## # weights:  15 (8 variable)
## initial  value 219.722458
## iter  10 value 179.982880
## final  value 179.981726
## converged
```

```r
# prog, program type: general academic, and vocation
# ses, social economic status, categorical
# write, continuous,
```

Multinomial Logit Model in R

Coefficients are log odds-ratios of the respective category over the reference category.

```
screenreg(m1, single.row = TRUE, beside = TRUE)
```

```
##
## =====================================================
##                   general           vocation
## -----------------------------------------------------
## (Intercept)        2.85 (1.17) *      5.22 (1.16) ***
## sesmiddle         -0.53 (0.44)        0.29 (0.48)
## seshigh           -1.16 (0.51) *     -0.98 (0.60)
## write             -0.06 (0.02) **    -0.11 (0.02) ***
## -----------------------------------------------------
## AIC              375.96             375.96
## BIC              402.35             402.35
## Log Likelihood  -179.98            -179.98
## Deviance         359.96             359.96
## Num. obs.        200                200
## K                  3                  3
## =====================================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

Multinomial Logit Interpretation

```
z <- summary(m1)$coefficients/summary(m1)$standard.errors
p <- (1 - pnorm(abs(z), 0, 1))*2
data.frame(p)
```

```
##           X.Intercept. sesmiddle    seshigh        write
## general   0.0144766100 0.2294379 0.02373856 6.818902e-03
## vocation  0.0000072993 0.5407530 0.09894976 3.176045e-07
```

Interpretation

```
data.frame(exp(coef(m1)))
```

```
##          X.Intercept. sesmiddle   seshigh     write
## general      17.32582 0.5866769 0.3126026 0.9437172
## vocation    184.61262 1.3382809 0.3743123 0.8926116
```

- All else being equal, a one unit increase in **writing** decreases the relative odds of being in **general program** over **academic program** by approximately 0.94 %, and decreases the relative odds of **vocational program** over **academic program** by 0.89 %.

Interpretation

- Again, we recommend, to use **predicted probabilities**.

```
head(fitted(m1))
```

```
##    academic   general  vocation
## 1 0.1482764 0.3382454 0.5134781
## 2 0.1202017 0.1806283 0.6991700
## 3 0.4186747 0.2368082 0.3445171
## 4 0.1726885 0.3508384 0.4764731
## 5 0.1001231 0.1689374 0.7309395
## 6 0.3533566 0.2377976 0.4088458
```

Interpretation

- Then, we create a scenario to plot the predicted probabilities.

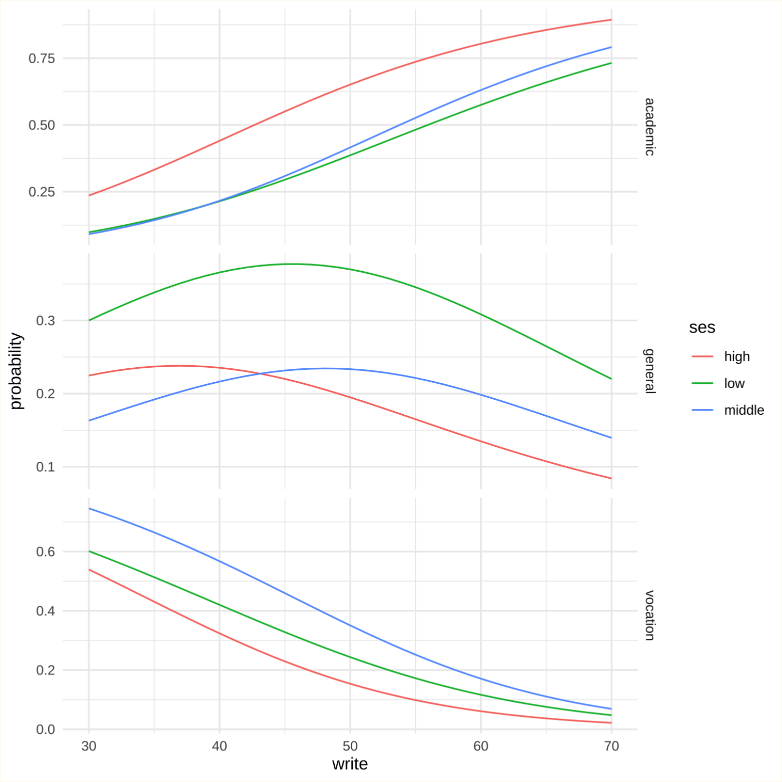
```
library(reshape2)
newdat <- data.frame(ses = rep(c("low", "middle", "high"),
                               each = 41),
                     write = rep(c(30:70), 3))

## store the predicted probabilities for each value of ses and write
pre <- cbind(newdat, predict(m1, newdata = newdat, type = "probs", se =
## melt data set to long for ggplot2
lpp <- melt(pre, id.vars = c("ses", "write"), value.name = "probability'
```

- Here, we are looking at the averaged predicted probabilities is to look at the averaged predicted probabilities for different values of the continuous predictor variable **write** within each level of **ses**.

Plot

```r
library(ggplot2)
ggplot(lpp, aes(x = write, y = probability, colour = ses)) +
  geom_line() + facet_grid(variable ~ ., scales = "free") +
  theme_minimal()
```



Independence of Irrelevant Alternatives (IIA)

- IIA assumption states that the addition of a new category does not change the odds-ratios of the existing alternatives.
- In other words, the alternatives need to be **"distinct and weighed independently in the eyes of each decision maker"**.

Test IIA assumption

- The Hausman-McFadden Test

```
library(mlogit)
#mlogit::hmftest(prog2 ~ 0 | ses + write, data = dat,
                #alt.subset = c("general",
                               #"vocation"))
```