

# Homework 3 Solution Script

Ling Zhu and Songeun Emily Lee

10/19/2025

## 0. Loading Packages

```
rm(list=ls()) # Start clean: remove all objects from the workspace
## Libraries
library(ggeffects)
library(foreign)
library(ordinal)
library(MASS)
library(car)
library(dplyr)
library(brant)
library(ggplot2)
library(stargazer)
library(nnet)
#library(MNP)
library(AER)
library(tidyverse)
library(lmtest)

## Working directory
setwd("/Users/songeunlee/Desktop/Ling/HW3")
#setwd("/Users/lingzhu/Dropbox/UH Teaching/POLS6382_2025/HW Assignments/2025 HW Review and Solution Scr

## Data
data1<-read.dta("Exercise3Part1.dta")
```

## 1. Analyzing Ordinal Data: Preliminaries

### Question 1a

```
# Descriptive Statistics: Dependent Variable (Abortion Opinions)

# (1) Frequency distribution (table)
freq_table <- table(data1$abortion)
freq_table
```

	1	2	3	4
	152	429	268	528

```
# (2) Proportion of each response category
prop.table(freq_table)
```

```
      1      2      3      4
0.1103849 0.3115468 0.1946260 0.3834423
```

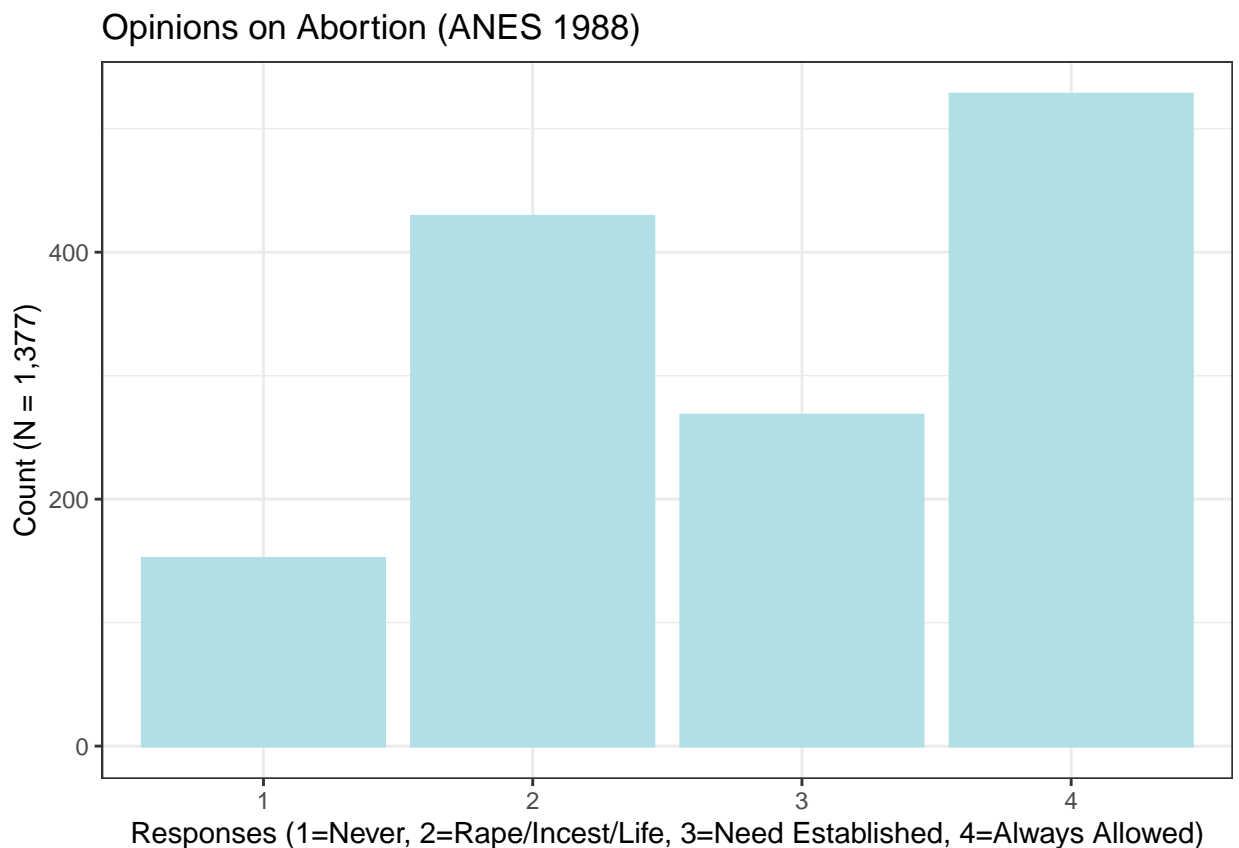
```
# (3) Mean and standard deviation (central tendency and dispersion)
mean(data1$abortion, na.rm = TRUE)
```

```
[1] 2.851126
```

```
sd(data1$abortion, na.rm = TRUE)
```

```
[1] 1.056018
```

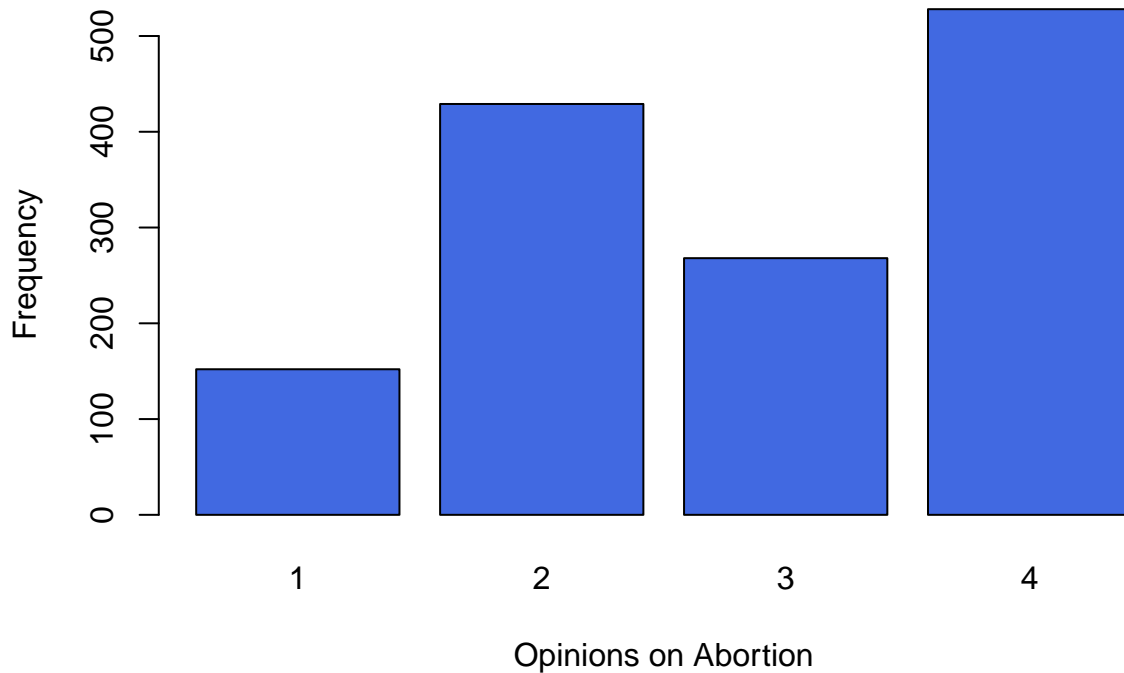
```
# (4) Visualization (ggplot bar chart)
ggplot(data1, aes(x = factor(abortion))) +
  geom_bar(fill = "powderblue", color = "powderblue") +
  labs(
    title = "Opinions on Abortion (ANES 1988)",
    x = "Responses (1=Never, 2=Rape/Incest/Life, 3=Need Established, 4=Always Allowed)",
    y = "Count (N = 1,377)"
  ) +
  theme_bw()
```



```
# (5) Simple base R barplot (for quick reference)
barplot(freq_table,
  main = "Q: Which of the opinions on abortion best agrees with your view?",
```

```
xlab = "Opinions on Abortion",
ylab = "Frequency",
col = "royalblue")
```

**Q: Which of the opinions on abortion best agrees with your view?**



```
# Descriptive Plots for Independent Variables (IVs)

# Arrange plots in 2x2 layout
par(mfrow = c(2, 2))

# (1) Political Ideology
# - Continuous 7-point scale: 1 = Extremely Liberal, 7 = Extremely Conservative
# - Histogram shows distribution across the ideological spectrum
iv1 <- hist(
  data1$conservatism,
  freq = TRUE,
  main = "Ideology",
  xlab = "Liberal-Conservative Ideology Scale (1-7)",
  xlim = c(0, 7),
  breaks = c(0, 1, 2, 3, 4, 5, 6, 7),
  ylim = c(10, 700),
  ylab = "Frequency",
  col = "steelblue"
)
text(iv1$mids, iv1$counts, labels = iv1$counts, adj = c(0.5, -0.5))

# (2) Education Level
# - Binary variable: 0 = High school or less; 1 = Attended college
college_f_freq <- table(data1$college)
iv2 <- barplot(
```

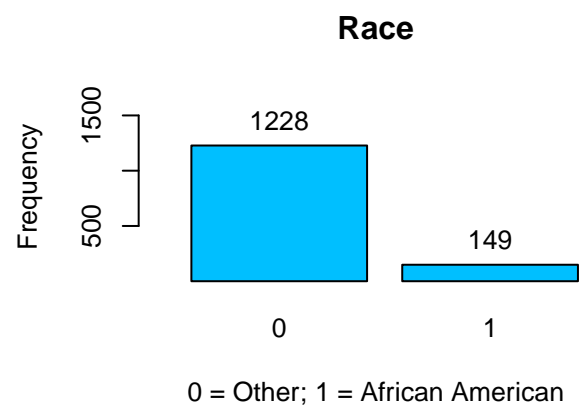
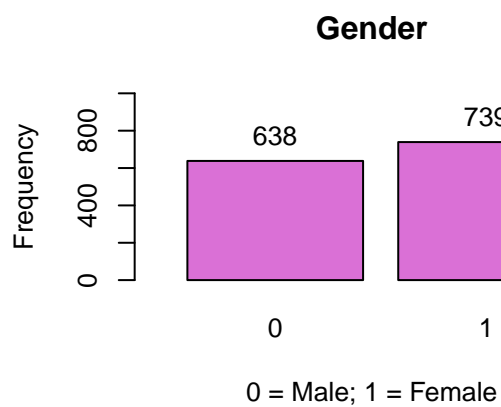
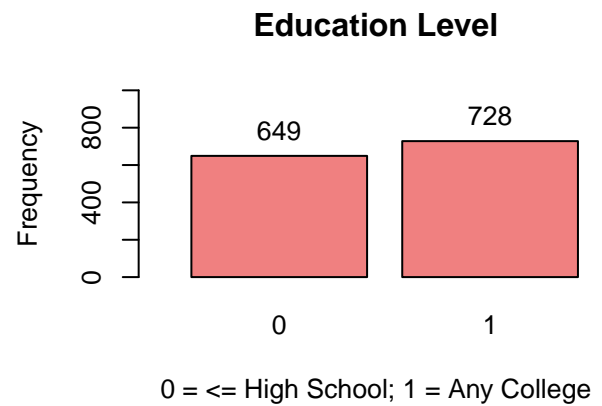
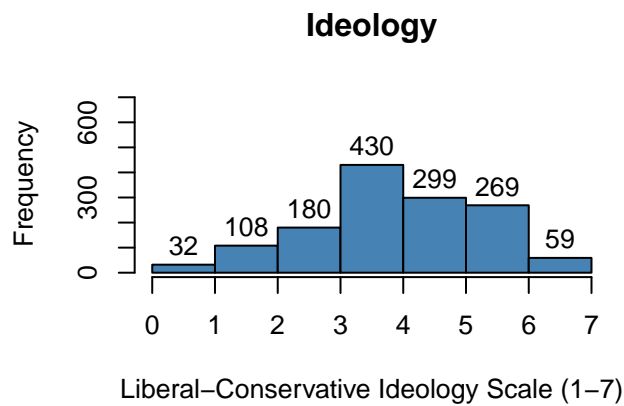
```

college_f_freq,
main = "Education Level",
xlab = "0 = <= High School; 1 = Any College",
ylab = "Frequency",
horiz = FALSE,
col = "lightcoral",
beside = FALSE,
xlim = c(0, 2.5),
ylim = c(0, 1000)
)
text(x = iv2, y = college_f_freq, labels = round(college_f_freq, digits = 3), pos = 3, col = "black")

# (3) Gender
# - Binary variable: 0 = Male; 1 = Female
female_f_freq <- table(data1$female)
iv3 <- barplot(
  female_f_freq,
  main = "Gender",
  xlab = "0 = Male; 1 = Female",
  ylab = "Frequency",
  horiz = FALSE,
  col = "orchid",
  beside = FALSE,
  xlim = c(0, 2.5),
  ylim = c(0, 1000)
)
text(x = iv3, y = female_f_freq, labels = round(female_f_freq, digits = 3), pos = 3, col = "black")

# (4) Race
# - Binary variable: 0 = Other; 1 = African American
black_f_freq <- table(data1$black)
iv4 <- barplot(
  black_f_freq,
  main = "Race",
  xlab = "0 = Other; 1 = African American",
  ylab = "Frequency",
  horiz = FALSE,
  col = "deepskyblue",
  beside = FALSE,
  xlim = c(0, 2.5),
  ylim = c(10, 1700)
)
text(x = iv4, y = black_f_freq, labels = round(black_f_freq, digits = 3), pos = 3, col = "black")

```



## Question 1b

```
# Model
model1<-polr(as.factor(abortion)~female+black+college+conservatism, data=data1,Hess=TRUE,model=TRUE)
summary(model1)
```

Call:

```
polr(formula = as.factor(abortion) ~ female + black + college +
      conservatism, data = data1, Hess = TRUE, model = TRUE)
```

Coefficients:

	Value	Std. Error	t value
female	-0.1002	0.09982	-1.0043
black	-0.1121	0.16421	-0.6828
college	0.6557	0.10082	6.5034
conservatism	-0.3205	0.03807	-8.4200

Intercepts:

	Value	Std. Error	t value
1 2	-3.3166	0.2120	-15.6418
2 3	-1.4531	0.1940	-7.4903
3 4	-0.6109	0.1907	-3.2041

Residual Deviance: 3449.791

AIC: 3463.791

```
# Odds Ratios
exp(coef(model1))
```

```
      female      black      college conservatism
0.9046164    0.8939337    1.9264304    0.7257675
```

```
(ci<-confint(model1))
```

```
      2.5 %      97.5 %
female    -0.2960048  0.09534168
black     -0.4333210  0.21098998
college    0.4584921  0.85377213
conservatism -0.3955819 -0.24630605
```

```
exp(cbind(OR = coef(model1), ci))
```

```
      OR      2.5 %      97.5 %
female    0.9046164  0.7437838  1.100035
black     0.8939337  0.6483524  1.234900
college    1.9264304  1.5816871  2.348489
conservatism 0.7257675 0.6732881 0.781683
```

## Question 1c: ggpredict

```
# Q1c - Predicted Probabilities with 95% CIs (using ggpredict)
```

```
library(MASS)      # polr()
library(ggeffects) # ggpredict()
library(ggplot2)   # plotting
library(ggpubr)    # arranging multiple plots
```

```
# --- Data hygiene: make sure DV is ordered factor ---
data1$abortion <- factor(data1$abortion, levels = c(1, 2, 3, 4), ordered = TRUE)
```

```
# --- 1) Baseline ordered logit model (same as Q1b) ---
```

```
# DV: abortion (ordered)
```

```
# IVs: female, black, college, conservatism
```

```
model1 <- polr(
  abortion ~ female + black + college + conservatism,
  data = data1,
  Hess = TRUE
)
```

```
# =====
```

```
# A. Predicted probabilities across the ideology scale
```

```
# - Shows how the probability of each abortion category changes
```

```
# as conservatism increases (holding other covariates at typical values)
```

```
# =====
```

```
# Generate predictions for ideology from 1 to 7 by 0.2 steps
```

```
pred_ideo <- ggpredict(model1, terms = c("conservatism[1:7 by=0.2]"))
```

```
# Plot: lines + CI ribbons by response category
```

```
fig_ideo <- ggplot(pred_ideo,
  aes(x = x, y = predicted,
```

```

        color = response.level, group = response.level)) +
geom_ribbon(aes(ymin = conf.low, ymax = conf.high, fill = response.level),
            alpha = 0.15, color = NA) +
geom_line(size = 0.9) +
scale_color_brewer(palette = "Dark2", name = "Abortion Attitudes",
                  labels = c("Complete Ban", "Conditional Ban",
                             "Conditional Permission", "Pro-Choice")) +
scale_fill_brewer(palette = "Dark2", guide = "none") +
labs(x = "Conservative Ideology (1-7)",
     y = "Predicted Probability",
     title = "Predicted Probabilities by Ideology") +
theme_bw() +
theme(legend.position = "bottom",
      axis.title = element_text(size = 13),
      plot.title = element_text(face = "bold"))

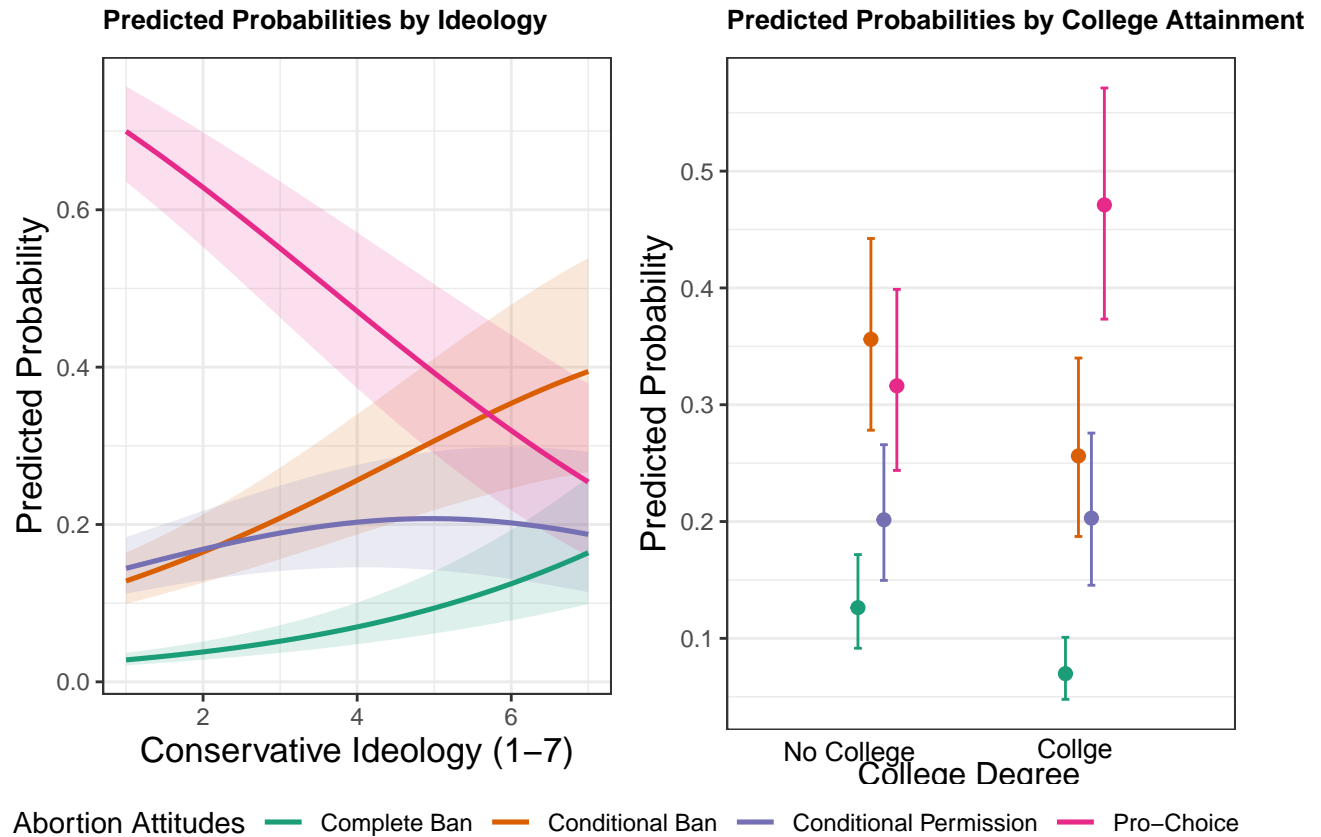
# =====
# B. Predicted probabilities by college attainment
#   - Compares probabilities for 0 (No College) vs 1 (College)
# =====

pred_college <- ggpredict(model1, terms = "college")

fig_college <- ggplot(pred_college,
                     aes(x = x, y = predicted,
                        color = response.level, group = response.level)) +
geom_point(position = position_dodge(width = 0.25), size = 2) +
geom_errorbar(aes(ymin = conf.low, ymax = conf.high),
              width = 0.15, position = position_dodge(width = 0.25)) +
scale_x_discrete(labels = c("0" = "No College", "1" = "College")) +
scale_color_brewer(palette = "Dark2", name = "Abortion Attitudes",
                  labels = c("Complete Ban", "Conditional Ban",
                             "Conditional Permission", "Pro-Choice")) +
labs(x = "College Degree",
     y = "Predicted Probability",
     title = "Predicted Probabilities by College Attainment") +
theme_bw() +
theme(legend.position = "bottom",
      axis.title = element_text(size = 13),
      plot.title = element_text(face = "bold"))

# Show A+B side-by-side, sharing one legend at bottom
ggarrange(fig_ideo, fig_college, nrow = 1, ncol = 2, common.legend = TRUE, legend = "bottom")

```



```
# =====
# C. Interaction model: conservatism × college
# - Estimates whether ideology's relationship with attitudes differs by college
# - Visualized as predicted probabilities across ideology, for each college group
# =====

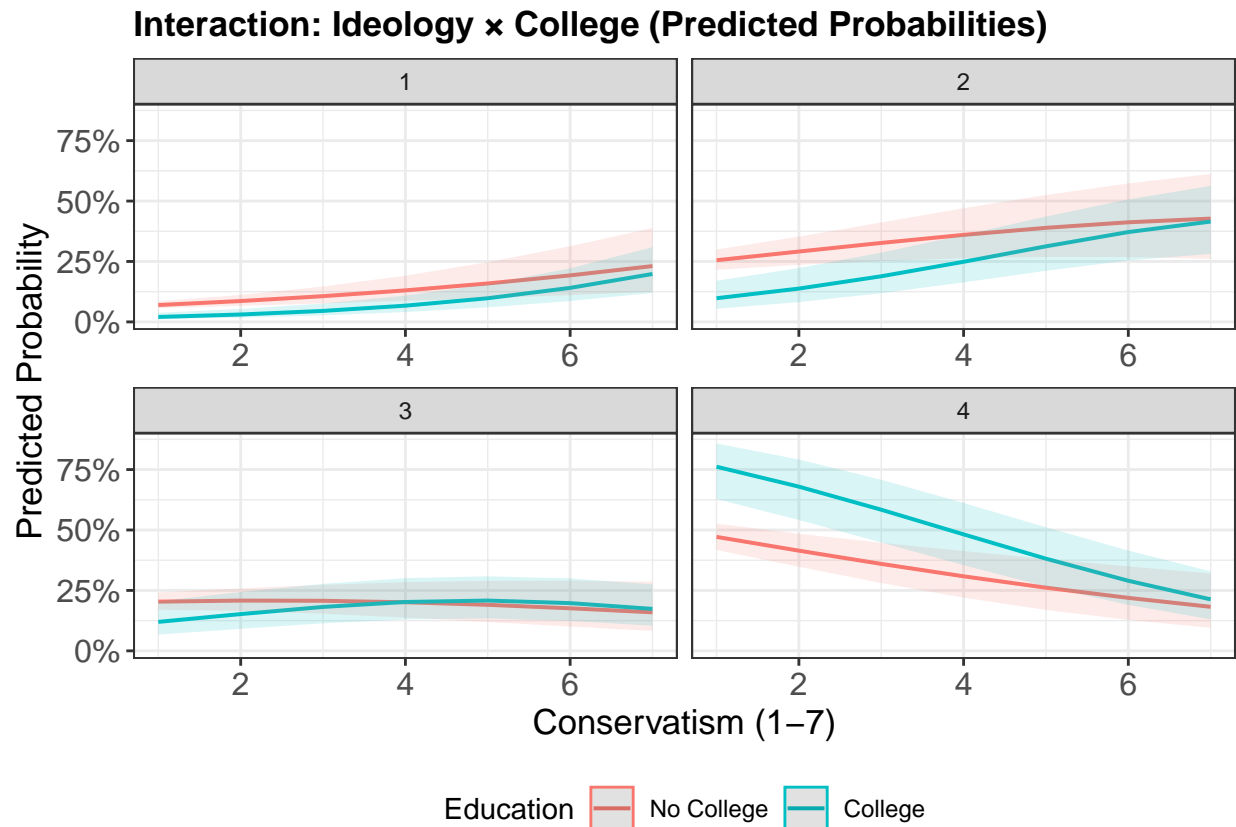
model_int <- polr(
  abortion ~ conservatism * college + black + female,
  data = data1,
  Hess = TRUE
)

# Predictions across ideology=1..7, separated by college = 0,1
pred_int <- ggpredict(model_int,
  terms = c("conservatism[1:7 by=1]", "college[0,1]"),
  ci.lvl = 0.95)

# ggpredict() returns a grouped data frame; `plot()` gives a good default.
# We add labels and theme for consistency.
fig_int <- plot(pred_int) +
  labs(x = "Conservatism (1-7)",
    y = "Predicted Probability",
    color = "Education",
    title = "Interaction: Ideology × College (Predicted Probabilities)") +
  scale_color_discrete(labels = c("No College", "College")) +
  theme_bw() +
  theme(legend.position = "bottom",
```

```
axis.text = element_text(size = 12),
axis.title = element_text(size = 13),
plot.title = element_text(face = "bold"))

# Show interaction plot
print(fig_int)
```



Question 1c: `ggeffect()`

```
# Predicted Probabilities with 95% CIs
library(ggeffects)
library(ggplot2)
library(ggpubr)

# Ideology
eff_ideo <- ggeffect(model1, terms = "conservatism [1:7 by=0.2]")
fig_ideo <- plot(eff_ideo) +
  labs(x = "Conservatism (1-7)", y = "Predicted Probability",
       title = "Predicted Probabilities by Ideology") +
  theme_bw() + theme(legend.position = "bottom", plot.title = element_text(face="bold"))

# College
eff_college <- ggeffect(model1, terms = "college")
if (!is.factor(eff_college$x)) {
  eff_college$x <- factor(eff_college$x, levels = c(0,1),
                        labels = c("No College", "College"))
}
```

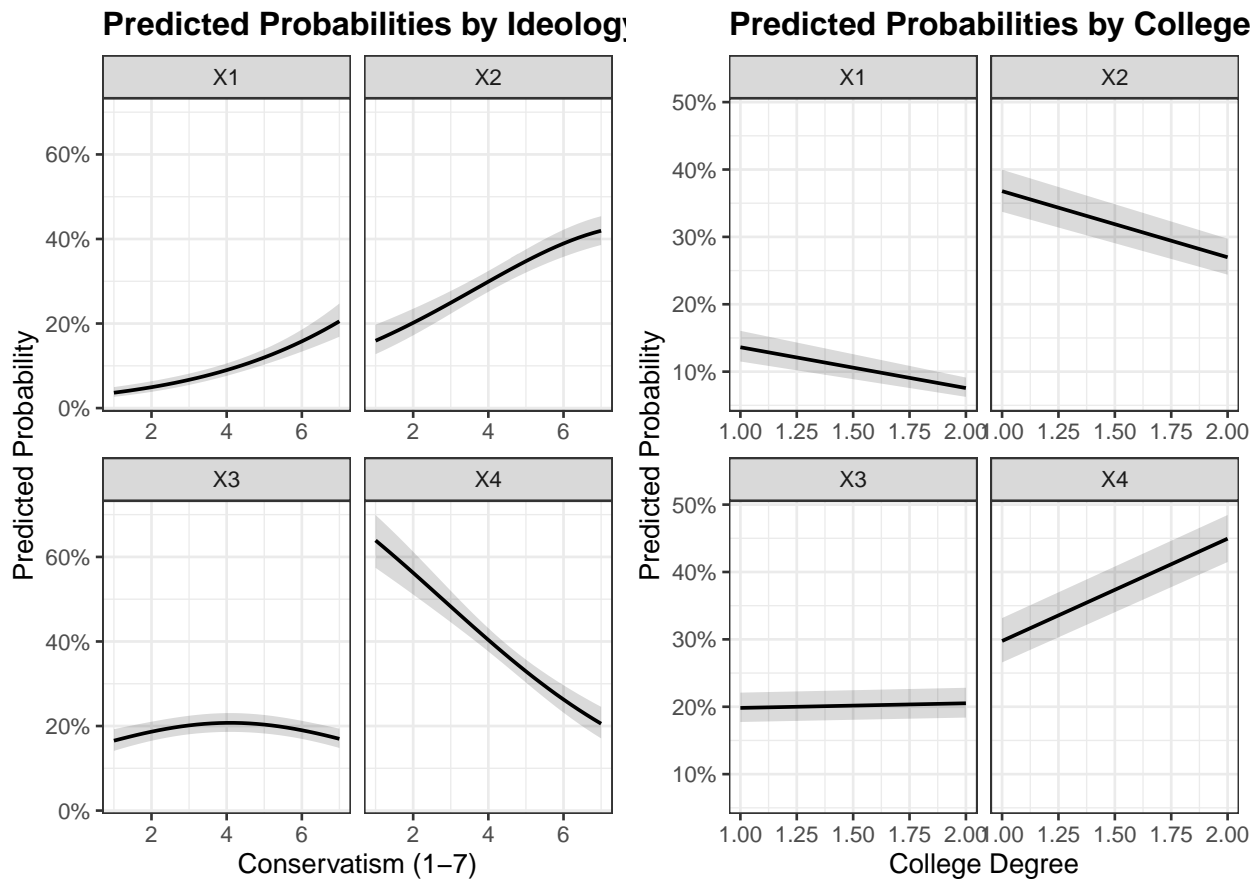
```

}
fig_college <- plot(eff_college) +
  labs(x = "College Degree", y = "Predicted Probability",
       title = "Predicted Probabilities by College") +
  theme_bw() + theme(legend.position = "bottom", plot.title = element_text(face="bold"))

# Interaction: conservatism × college (
eff_int <- ggeffect(model_int, terms = c("conservatism [1:7]", "college"))
if (!is.null(eff_int$group) && !is.factor(eff_int$group)) {
  eff_int$group <- factor(eff_int$group, levels = c(0,1),
                        labels = c("No College","College"))
}
fig_int <- plot(eff_int) +
  labs(x = "Conservatism (1-7)", y = "Predicted Probability",
       color = "Education", fill = "Education",
       title = "Interaction: Ideology × College") +
  theme_bw() + theme(legend.position = "bottom", plot.title = element_text(face="bold"))

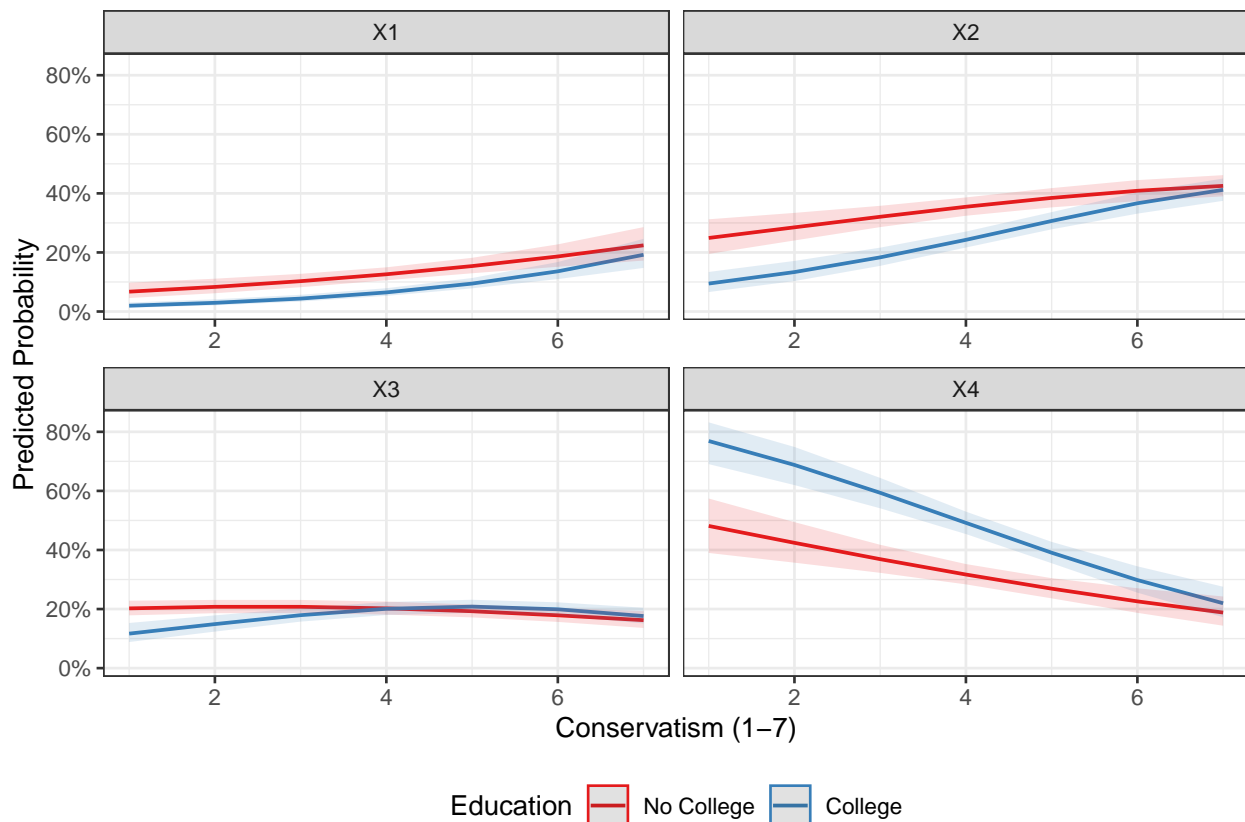
ggarrange(fig_ideo, fig_college, nrow = 1, ncol = 2, common.legend = TRUE, legend = "bottom")

```



```
print(fig_int)
```

### Interaction: Ideology × College



### Question 1d

```
# Formal Likelihood-Ratio Test
# Compare model fits
dev1 <- deviance(model1)
aic1 <- AIC(model1)

# Stepwise model selection
fit <- step(model1)
```

Start: AIC=3463.79  
 abortion ~ female + black + college + conservatism

	Df	AIC
- black	1	3462.3
- female	1	3462.8
<none>		3463.8
- college	1	3504.6
- conservatism	1	3535.1

Step: AIC=3462.26  
 abortion ~ female + college + conservatism

	Df	AIC
- female	1	3461.3
<none>		3462.3

```
- college      1 3503.3
- conservatism 1 3533.4
```

Step: AIC=3461.35

abortion ~ college + conservatism

```
      Df    AIC
<none>      3461.3
- college    1 3503.0
- conservatism 1 3532.0
```

```
dev2 <- deviance(fit)
aic2 <- AIC(fit)

# Compare deviance difference (nested model test)
diff_dev <- dev2 - dev1
df_diff  <- model1$edf - fit$edf
p_val    <- pchisq(diff_dev, df_diff, lower.tail = FALSE)
```

```
## Overall model fit by comparing more than one model
## Looking at how many variables are significant and their magnitude
deviance(model1)
```

```
[1] 3449.791
```

```
AIC(model1)
```

```
[1] 3463.791
```

```
fit <- step(model1) # Lowest AIC indicates the best model
```

Start: AIC=3463.79

abortion ~ female + black + college + conservatism

```
      Df    AIC
- black      1 3462.3
- female      1 3462.8
<none>        3463.8
- college     1 3504.6
- conservatism 1 3535.1
```

Step: AIC=3462.26

abortion ~ female + college + conservatism

```
      Df    AIC
- female      1 3461.3
<none>        3462.3
- college     1 3503.3
- conservatism 1 3533.4
```

Step: AIC=3461.35

abortion ~ college + conservatism

```
      Df    AIC
<none>      3461.3
- college    1 3503.0
- conservatism 1 3532.0
```

```
deviance(fit)
```

```
[1] 3451.348
```

```
deviance(fit) - deviance(model1)
```

```
[1] 1.556917
```

```
pchisq(1.55, model1$edf - fit$edf, lower = F)
```

```
[1] 0.4607038
```

## Question 1e

```
# One way to complete
```

```
fm1<-clm(as.factor(abortion)~female+black+college+conservatism, data=data1)
summary(fm1)
```

```
formula: as.factor(abortion) ~ female + black + college + conservatism
data:    data1
```

```
link threshold nobs logLik AIC niter max.grad cond.H
logit flexible 1377 -1724.90 3463.79 5(0) 2.87e-10 1.1e+03
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
female	-0.10024	0.09982	-1.004	0.315
black	-0.11213	0.16421	-0.683	0.495
college	0.65567	0.10082	6.503	7.85e-11 ***
conservatism	-0.32053	0.03807	-8.420	< 2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Threshold coefficients:

	Estimate	Std. Error	z value
1 2	-3.3167	0.2120	-15.642
2 3	-1.4531	0.1940	-7.490
3 4	-0.6109	0.1907	-3.204

```
con.nom<-clm(as.factor(abortion)~female+black+college, nominal=~conservatism, data=data1)
anova(fm1,con.nom)
```

Likelihood ratio tests of cumulative link models:

```
formula:
fm1      as.factor(abortion) ~ female + black + college + conservatism
con.nom  as.factor(abortion) ~ female + black + college
nominal:      link: threshold:
fm1      ~1          logit flexible
con.nom  ~conservatism logit flexible
```

	no.par	AIC	logLik	LR.stat	df	Pr(>Chisq)
fm1	7	3463.8	-1724.9			
con.nom	9	3465.2	-1723.6	2.6338	2	0.268

```
black.nom<-clm(as.factor(abortion)~female+college+conservatism, nominal=~black, data=data1)
anova(fm1,black.nom)
```

Likelihood ratio tests of cumulative link models:

```
      formula:
fm1      as.factor(abortion) ~ female + black + college + conservatism
black.nom as.factor(abortion) ~ female + college + conservatism
      nominal: link: threshold:
fm1      ~1      logit flexible
black.nom ~black  logit flexible

      no.par    AIC  logLik LR.stat df Pr(>Chisq)
fm1      7 3463.8 -1724.9
black.nom 9 3466.3 -1724.2  1.4442  2      0.4857
female.nom<-clm(as.factor(abortion)~black+college+conservatism, nominal=~female, data=data1)
anova(fm1,female.nom)
```

Likelihood ratio tests of cumulative link models:

```
      formula:
fm1      as.factor(abortion) ~ female + black + college + conservatism
female.nom as.factor(abortion) ~ black + college + conservatism
      nominal: link: threshold:
fm1      ~1      logit flexible
female.nom ~female logit flexible

      no.par    AIC  logLik LR.stat df Pr(>Chisq)
fm1      7 3463.8 -1724.9
female.nom 9 3455.9 -1718.9  11.914  2      0.002588 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

coll.nom<-clm(as.factor(abortion)~black+female+conservatism, nominal=~college, data=data1)
anova(fm1,coll.nom)
```

Likelihood ratio tests of cumulative link models:

```
      formula:                                     nominal:
fm1      as.factor(abortion) ~ female + black + college + conservatism ~1
coll.nom as.factor(abortion) ~ black + female + conservatism      ~college
      link: threshold:
fm1      logit flexible
coll.nom logit flexible

      no.par    AIC  logLik LR.stat df Pr(>Chisq)
fm1      7 3463.8 -1724.9
coll.nom 9 3461.6 -1721.8   6.21  2      0.04483 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# --- Brant Test for the Proportional Odds Assumption ---
# Used after fitting an ordered logit model to check whether
# each predictor has a constant effect across outcome categories.
#
# H0: The proportional-odds assumption holds.
```

```
# H1: At least one variable violates this assumption.
#
# Interpretation:
# - If the Omnibus p-value is significant ( $p < 0.05$ ), it indicates that
#   the proportional-odds assumption is violated for the model as a whole.
# - If a variable's p-value is significant ( $p < 0.05$ ), it indicates that
#   the proportional-odds assumption is violated for that specific variable.
```

```
brant(model1)
```

```
-----
Test for      X2  df  probability
-----
Omnibus       18.47   8    0.02
female        10.33   2    0.01
black          0.68   2    0.71
college        4.15   2    0.13
conservatism   1.61   2    0.45
-----
```

H0: Parallel Regression Assumption holds

```
#You could also use the clm model or the "crude way" using prop. table
```

## Question 2a

```
data2<-read.dta("Exercise3Part2.dta")
attach(data2)

# Display summary statistics (median values) for key variables
stargazer(data2[c("prezvote", "respideo", "democrat", "GOP")], median = T,
           type = "text", digits = 2, title = "Descriptive Statistics")
```

### Descriptive Statistics

```
=====
Statistic  N  Mean St. Dev. Min Median Max
-----
prezvote   351 1.60   0.66   1    2    3
respideo   351 4.31   1.40   1    4    7
democrat   351 0.37   0.48   0    0    1
GOP         351 0.38   0.49   0    0    1
-----
```

## Question 2b

```
# Recode 'prezvote' as a labeled factor (categorical variable)
data2$prezvote <- factor(data2$prezvote, labels=c("Bill Clinton", "Robert Dole", "H. Ross Perot"))

# Estimate multinomial logistic regression
# Dependent variable: prezvote (3 categories)
# Independent variables: respideo (ideology), democrat, GOP
multinomialmod <- multinom(prezvote ~ respideo + democrat + GOP, data2)
```

```
# weights: 15 (8 variable)
initial value 385.612913
iter 10 value 190.739234
final value 190.558504
converged

# Regression Results
stargazer(multinomialmod, type="text",
          title = "The multinomial regression of presidential voting in 1996")
```

The multinomial regression of presidential voting in 1996

```
=====
Dependent variable:
-----
Robert Dole    H. Ross Perot
(1)           (2)
-----
respieo       1.255***    0.487***
              (0.203)    (0.187)

democrat      -3.640***    -1.647***
              (0.701)    (0.459)

GOP           1.118***    -0.456
              (0.404)    (0.561)

Constant      -5.547***    -2.734***
              (0.910)    (0.795)

-----
Akaike Inf. Crit.    397.117    397.117
=====
Note:                *p<0.1; **p<0.05; ***p<0.01
```

```
## Interpretation based on odds
exp(coef(multinomialmod))
```

```
(Intercept) respieo  democrat    GOP
Robert Dole  0.003899364 3.509456 0.02625963 3.0586143
H. Ross Perot 0.064947141 1.627988 0.19271795 0.6339505
```

## Question 2c and 2d

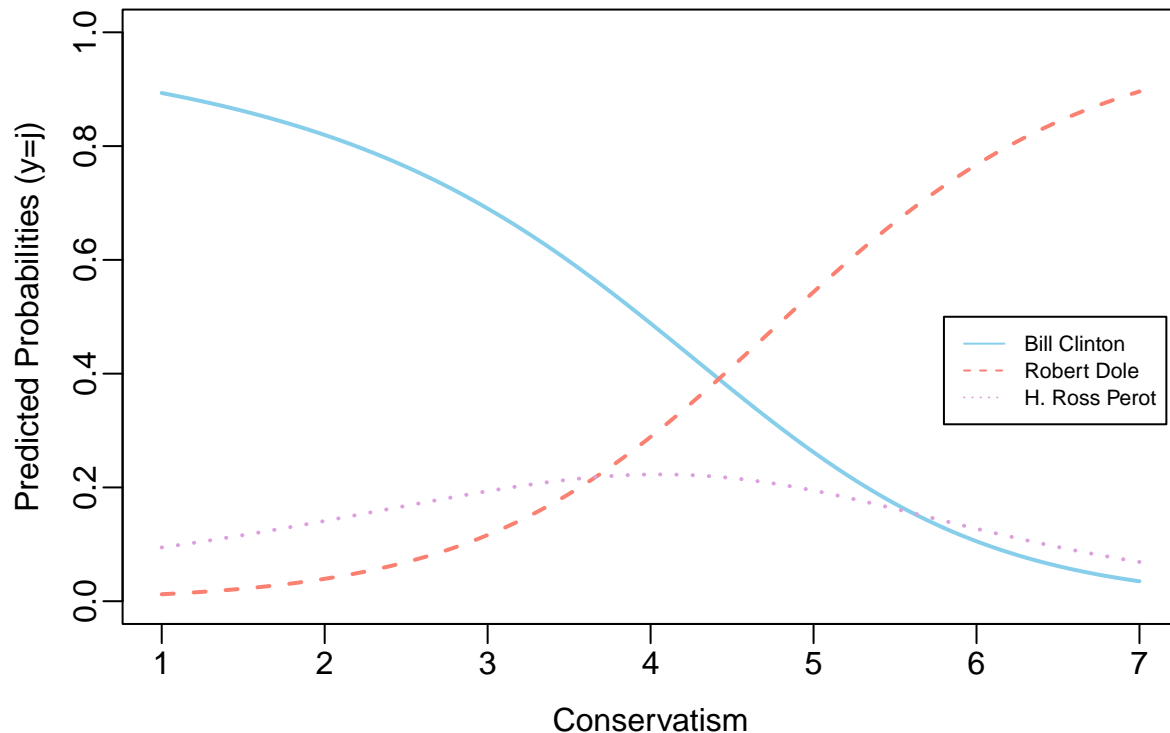
```
## For conservatism:
# Create a sequence of conservatism values (1 to 7) while holding
# democrat and GOP at their median values (to isolate ideology effect)
predata_conserv <- cbind(
  respieo = seq(1, 7, length=500),
  democrat = median(democrat, na.rm=T),
  GOP = median(GOP, na.rm=T))

# Predict the probability of voting for each candidate across the conservatism scale
prob_conserv <- predict(multinomialmod, predata_conserv, type="probs")
```

```
## For democrats:
# When the respondent identifies as Democrat (democrat = 1, GOP = 0)
# and ideology is held at its mean
predata_dem <- cbind(respideo = mean(respideo, na.rm = T), democrat = 1, GOP = 0)
prob_dem <- predict(multinomialmod, predata_dem, type="probs")

## For GOP:
# When the respondent identifies as Republican (democrat = 0, GOP = 1)
predata_gop <- cbind(respideo = mean(respideo, na.rm = T), democrat = 0, GOP = 1)
prob_gop <- predict(multinomialmod, predata_gop, type="probs")

## Plotting the probabilities against conservatism
par(mfrow = c(1,1), mar=c(3.5,4,1,1), oma=c(0,0,2,0), mgp=c(2,0.5,0))
plot(c(1,7), c(0,1), type='n',
     xlab="Conservatism",
     ylab="Predicted Probabilities (y=j)",
     main = "")
lines(seq(1, 7, length=500), prob_conserv[,1], lty=1, lwd=2, col="skyblue")
lines(seq(1, 7, length=500), prob_conserv[,2], lty=2, lwd=2, col="salmon")
lines(seq(1, 7, length=500), prob_conserv[,3], lty=3, lwd=2, col="plum")
legend(5.8, 0.5, cex=0.7, lty=1:3,
     c("Bill Clinton", "Robert Dole", "H. Ross Perot"),
     col=c("skyblue", "salmon", "plum"))
```



```
## Show the probabilities against different partisan identification:
partisanship <- rbind(prob_dem, prob_gop); partisanship
```

	Bill Clinton	Robert Dole	H. Ross Perot
prob_dem	0.8887121	0.02038555	0.09090231
prob_gop	0.2494867	0.66656842	0.08394491

## Question 2e

```
# Recode the dependent variable (prezvote) so that "H. Ross Perot" becomes the new baseline category
# This means all coefficients will now be interpreted relative to Perot.
data2$recodedprezvote <- relevel(data2$prezvote, ref = "H. Ross Perot")

# Estimate a new multinomial logistic regression using the new reference category
# The predictors (respideo, democrat, GOP) remain the same.
multinomialmod2 <- multinom(recodedprezvote ~ respideo + democrat + GOP, data2)

# weights:  15 (8 variable)
initial  value 385.612913
iter   10 value 190.955278
final   value 190.558504
converged

# Compare the two multinomial models side by side:
# - multinomialmod : baseline = Bill Clinton (default)
# - multinomialmod2 : baseline = H. Ross Perot (relevelled)
# The goal is to see how changing the baseline affects coefficient signs and magnitudes.
stargazer(multinomialmod, multinomialmod2,
           type = "text",
           title = "The comparison of multinomial models based on different baseline")
```

The comparison of multinomial models based on different baseline

Dependent variable:				
	Robert Dole (1)	H. Ross Perot (2)	Bill Clinton (3)	Robert Dole (4)
respideo	1.255*** (0.203)	0.487*** (0.187)	-0.487*** (0.187)	0.768*** (0.224)
democrat	-3.640*** (0.701)	-1.647*** (0.459)	1.647*** (0.459)	-1.993*** (0.770)
GOP	1.118*** (0.404)	-0.456 (0.561)	0.456 (0.561)	1.574*** (0.525)
Constant	-5.547*** (0.910)	-2.734*** (0.795)	2.734*** (0.795)	-2.813*** (1.026)
Akaike Inf. Crit.	397.117	397.117	397.117	397.117

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

After re-specifying the baseline category from Bill Clinton to H. Ross Perot, the direction and magnitude of the coefficients change as expected, but the substantive interpretation remains the same. In a multinomial logistic regression, changing the baseline simply alters which candidate serves as the comparison group—it does not affect the underlying relationships among variables.

When Perot is the new baseline, the coefficients for Clinton and Dole switch signs relative to their previous positions. For instance, the coefficient for ideology (respideo) becomes -0.49 for Clinton and +0.77 for Dole,

indicating that as respondents become more conservative, they are less likely to vote for Clinton and more likely to vote for Dole compared to Perot. Similarly, Democratic identification strongly increases support for Clinton and decreases support for Dole, while Republican identification sharply raises the likelihood of choosing Dole over either Clinton or Perot.

Overall, this re-specification does not reveal new substantive information—it simply reframes the comparison. The same ideological and partisan patterns remain: conservatives and Republicans prefer Dole, liberals and Democrats prefer Clinton, and Perot continues to attract moderate, less partisan voters. Changing the baseline helps confirm the internal consistency of the model but does not alter its substantive conclusions.

## Question 2F

```
require(statmod)
require(remotes)
require(mlogit)

# To learn how hmftest() works, try ?hmftest or example(hmftest)
# help("hmftest")
# example(hmftest)

# Keep only needed variables
newdata <- subset(data2, select = c("prezvote", "respideo", "democrat", "GOP"))

# Convert to mlogit format
iiadata <- mlogit.data(newdata, shape = "wide", choice = "prezvote")

# Multinomial logit with INDIVIDUAL-SPECIFIC regressors only (no ASCs)
mlogitmod1 <- mlogit(prezvote~0|respideo+democrat+GOP, data=iiadata)

# Refit on a RESTRICTED CHOICE SET (drop Clinton)
mlogitmod2 <- mlogit(prezvote~0|respideo+democrat+GOP,
                    data=iiadata,
                    alt.subset=c("Robert Dole", "H. Ross Perot"))

# Hausman-McFadden test for IIA (full model vs. restricted model)
hmftest(mlogitmod1, mlogitmod2)
```

Hausman-McFadden test

```
data: iiadata
chisq = 42.707, df = 4, p-value = 1.191e-08
alternative hypothesis: IIA is rejected
```

The Independence of Irrelevant Alternatives (IIA) assumption in multinomial logit models states that the relative odds of choosing one alternative over another are independent of the presence or characteristics of other options. In other words, removing or adding a third candidate should not affect the relative probability between the remaining two.

To evaluate this assumption, I conducted a Hausman–McFadden test `hmftest` by comparing the full model (including Clinton, Dole, and Perot) with a restricted model excluding Bill Clinton. The test statistic is 42.71 with 4 degrees of freedom and a p-value less than 0.001. Since the p-value is highly significant, I reject the IIA assumption.

Statistically, this means that the choice probabilities are not independent across alternatives—% respondents' preferences among Dole, Perot, and Clinton are correlated. Substantively, this suggests that voters may view

the three candidates as imperfect substitutes; for example, support for Perot is likely related to attitudes toward Dole rather than being an independent third option. Therefore, the standard multinomial logit may not fully capture the interdependence among choices, and more flexible models such as the nested logit could provide a better fit.