# Ordered and Multinomial Logit

## Implementation and interpretation

Dominique Lockett

Washington University in St. Louis

# Introduction

# Table of contents

# Introduction Logistic Regression

- Limited outcomes in the dependent variable
- Use logarithmic transformation on the outcome variable
    - model a nonlinear association in a linear way

# Ordered Logit

# Motivation Ordered

- Finite and discrete values with more than two outcomes

# Motivation Ordered

- Finite and discrete values with more than two outcomes
- Data with meaningful sequential values
  - income levels
    (0,10000],(10000,30000],(30000, $\infty$]
  - Likert-type scale

# Motivation Ordered

- Finite and discrete values with more than two outcomes
- Data with meaningful sequential values
  - income levels
    (0,10000],(10000,30000],(30000, $\infty$]
  - Likert-type scale
  - ~~Gender~~
  - ~~Party affiliation~~
  - ~~Education~~

# Mechanics Ordered

**Proportional Odds Assumption[2]**

The assumption that the explanatory variables have the same effect on the odds regardless of the threshold.

poor, $\log \frac{p_1}{p_2+p_3+p_4+p_5}$, 0

poor or fair, $\log \frac{p_1+p_2}{p_3+p_4+p_5}$, 1

poor,fair, or good,, $\log \frac{p_1+p_2+p_3}{p_4+p_5}$, 2

poor, fair, good, or very good, $\log \frac{p_1+p_2+p_3+p_4}{p_5}$, 3

# Math Ordered

## Formula

$$y_i^* = \alpha + \mathbf{x}_i'\beta + \epsilon_i = \alpha + Z_i + \epsilon_i$$

$y_i^*$ = latent utility
where $\epsilon_i$ Logistic(0,$\sigma^2$)

## Distribution

$$F(Z_i) = exp(Z_i)/1 + exp(Z_i)$$

$$y_i = 1, \quad if \quad Z_i \leq \eta_1 \quad \text{(Region 1)}$$
$$y_i = j, \quad if \quad \eta_{j-1} < Z_i \leq \eta_j \quad \text{(Region j)}$$
$$y_i = J, \quad if \quad \eta_{J-1} < Z_i \quad \text{(Region J)}$$

where $\eta_1 < \eta_2 < \eta_3, .., \eta_n$ & $\eta_1 \geq 0$,
parameters known as *thresholds* or *cutpoints*

# Math Ordered

## Marginal Effects:

Obtained by evaluating the appropriate density functions at the relevant points and multiplying by the associated coefficient [1]

## Continuous:

$$\frac{d}{dx}\left[\frac{\exp x}{1+\exp x}\right]$$

$$= \frac{[1+\exp(x)]\exp(x) - [\exp(x)]^2}{[1+\exp(x)]^2}$$

# Math Intuition Ordered

$$logit[P(Y \leq j)] = \alpha + \mathbf{x}_i'\beta, j - 1, ...J - 1$$

Prob. answering specific level consv. given party

| | Democrat[1] | Republican[0] |
|---|---|---|
| Very Liberal[1] | 0.1832505 | 0.07806044 |
| Slightly Liberal[2] | 0.1942837 | 0.10819225 |
| Moderate[3] | 0.3930552 | 0.37275214 |
| Slightly Conservative[4] | 0.1147559 | 0.18550357 |
| Very Conservative[5] | 0.1146547 | 0.25549160 |

Call:
polr(formula=pol.ideology party, data = dat

Coefficients:

| | Value | Std. Error | t-value |
|---|---|---|---|
| partyDem | -0.9745 | 0.1292 | -7.545 |
| VeryLiberal\|Slightly Liberal | -2.4690 | 0.1318 | -18.7363 |
| Slightly Liberal\|Moderate | -1.4745 | 0.1090 | -13.5314 |
| Moderate\|Slightly Conservative | 0.2371 | 0.0942 | 2.5165 |
| Slightly Conservative\|Very Conservative | 1.0695 | 0.1039 | 10.2923 |

If we wanted to find the odds a Democrat identifies as 'Slightly Liberal' or less:

$$\text{logit}[\text{P}(\text{Y} \leq 2)] = -1.4745 - -0.9745(1) = -0.5$$

$$P(Y \leq j) = \frac{exp(\alpha + \mathbf{x}_i'\beta)}{1 + exp(\alpha + \mathbf{x}_i'\beta)} \Rightarrow \frac{exp(-0.5)}{1 + exp(-0.5)} = .378$$

# Math Intuition Ordered

$$logit[P(Y \leq j)] = \alpha + \mathbf{x}_i'\beta, j - 1, ...J - 1$$

Prob. answering specific level consv. given party

| | Democrat [1] | Republican [0] |
|---|---|---|
| Very Liberal [1] | 0.1832505 | 0.07806044 |
| Slightly Liberal [2] | 0.1942837 | 0.10819225 |
| Moderate [3] | 0.3930552 | 0.37275214 |
| Slightly Conservative [4] | 0.1147559 | 0.18550357 |
| Very Conservative [5] | 0.1146547 | 0.25549160 |

Call:
polr(formula=pol.ideology party, data = dat

Coefficients:

| | Value | Std. Error | t-value |
|---|---|---|---|
| partyDem | -0.9745 | 0.1292 | -7.545 |
| VeryLiberal\|Slightly Liberal | -2.4690 | 0.1318 | -18.7363 |
| Slightly Liberal\|Moderate | -1.4745 | 0.1090 | -13.5314 |
| Moderate\|Slightly Conservative | 0.2371 | 0.0942 | 2.5165 |
| Slightly Conservative\|Very Conservative | 1.0695 | 0.1039 | 10.2923 |

If we wanted to find the odds a Democrat identifies as 'Slightly Liberal' or less:

$$\text{logit}[\text{P}(\text{Y} \leq 2)] = -1.4745 - -0.9745(1) = -0.5$$

$$P(Y \leq j) = \frac{exp(\alpha + \mathbf{x}_i'\beta)}{1 + exp(\alpha + \mathbf{x}_i'\beta)} \Rightarrow \frac{exp(-0.5)}{1 + exp(-0.5)} = .378$$

# Benefits Ordered

- Non-linear

# Benefits Ordered

- Non-linear
- Overcomes OLS i.i.d. assumption

# Benefits Ordered

- Non-linear
- Overcomes OLS i.i.d. assumption
- Applicable to discrete or continuous independent variables

# Shortcomings Ordered

- Vague
  - Move toward larger values of dv

# Shortcomings Ordered

- Vague
  - Move toward larger values of dv
- Trivial differences
  - Quasi-normal data w/ 3-4 scale dv

# Multinomial Logit

# Motivation Multinomial

- Discrete, mutually exclusive, unordered dependent variables
    - Party ID
        - 0=Republican, 1=Independent, 2=Democrat

# Motivation Multinomial

- Discrete, mutually exclusive, unordered dependent variables
    - Party ID
        - 0=Republican, 1=Independent, 2=Democrat
    - Region

# Motivation Multinomial

- Discrete, mutually exclusive, unordered dependent variables
  - Party ID
    - 0=Republican, 1=Independent, 2=Democrat
  - Region
- Choice Model

# Mechanics Multinomial

- $M$ outcomes
  - $M$-1 binary logistic regression models
- Extension of binomial logistic regression
- Probability of success in a given category $M$

# Mechanics Multinomial

**Independence of Irrelevant Alternatives (IIA):**

The **assumption** that the introduction or improvement of any alternative will have the same proportional impact on the original alternatives

**Example**

1= Train ; 2= Bus ; 3 = Car
IIA assumes that adding a $4^{th}$ option, a bike, will not have any impact on the probability of choosing your original 1-3 choices

# Math Multinomial

Formula

$$Z_{ij} = \sum_{r=1}^{R} \beta_{jr} X_{ir}$$

Normalization

$$Pr(Y = 1) = \frac{1}{1+\sum_{j=2}^{M} \exp(Z_{ij})}$$

$$Pr(Y_i = K) = \frac{\exp(Z_{ik})}{1+\sum_{j=2}^{M} \exp(Z_{ij})}$$

# Math Multinomial

**Risk Ratio[3]:**

The logarithm of the ratio of the probability of outcome m to that of outcome k

k = excluded observation

$$\left(\frac{Pr(Y_i\!=\!m)}{Pr(Y_i\!=\!1)}\right) = \exp\left(Z_{im}\right)$$

# Benefits Multinomial

- Does not assume
  - normality
  - linearity
  - homoscedasticity

# Benefits Multinomial

- Does not assume
  - normality
  - linearity
  - homoscedasticity
- Indepenedent variables
  - can be unbounded
  - needn't be interval

## Shortcomings Multinomial

- IIA assumption

# Examples

# TAPS Data Multinomial

- pid
  - Generally speaking, do you usually think of yourself as [Republican/a Democrat] and independent?
    - 1-Democrat; 2-Independent; 3-Republican
- abort
  - Do you generally support or oppose a woman's right to abortion
    - 1-support; 2-oppose;
- taxes
  - Please tell me if you would favor or oppose a federal tax policy that increases income taxes for people with the highest incomes .
    - 1-support; 2-oppose;
- n = 1164

# Example Multinomial

```
#library(nnet)
> m2 <- multinom(pid ~ abort + taxes, data = dat2)
> summary(m2)
```

```
Coefficients:
  (Intercept)     abort    taxes
2   -3.636725 0.8829103 2.034146
3   -6.683510 1.7884000 3.316103
```

# Example Multinomial

```
>Anova(m2)
```

---

```
Analysis of Deviance Table (Type II tests)

Response: pid
      LR Chisq Df Pr(>Chisq)
abort    75.68  2  < 2.2e-16 ***
taxes   231.17  2  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
Residual Deviance: 2086.295
AIC: 2098.295
```

## Example Multinomial

Risk Ratio:

```
> exp(coef(m1))
```

```
  (Intercept)     abort     taxes
2 0.026338453  2.417926  7.645719
3 0.001251378  5.979877 27.552755
```

# TAPS Data Ordered

- y
  - Do you agree or disagree that an American-born child of illegal immigrants should not be considered a U.S. citizen.
    - 1-strongly oppose; 2-oppose; 3-neutral; 4-support; 5-strongly support
- dt_strong
  - In your opinion, how well does the phrase ’is a strong leader’ describe Donald Trump?
    - 1-not well at all; 2-slightly well; 3-moderately well; 4-very well; 5-extremely well
- party_id
  - Generally speaking, do you usually think of yourself as [Republican/a Democrat] and independent?
    - 1-Democrat; 2-Independent; 3-Republican
- n = 1332

# Example Ordered

```
> m1 <- polr(y ~ dt_strong + party_id, data = dat)
> summary(m1)
```

```
Coefficients:
           Value  Std. Error  t value
dt_strong  0.2924    0.03800    7.693
party_id   0.6589    0.06776    9.723
```

# Example Ordered

```
Intercepts:
                     Value   Std. Error t value
s. oppose|oppose     0.4459  0.1329      3.3550
oppose|nuetral       1.7214  0.1388     12.4035
nuetral|support      2.3707  0.1466     16.1738
support|s. support   3.5660  0.1640     21.7435

Residual Deviance: 3983.457
AIC: 3995.457
```

Residual Deviance: 717.0249 AIC: 727.0249

# Example Ordered

```
> pval <- Anova(m1)
> pval
```

```
Response: y
          LR Chisq  Df  Pr(>Chisq)
dt_strong   59.977   1  9.595e-15 ***
party_id    96.387   1  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

# Example Ordered

Log Odds:

```
> exp(coef(m1))
```

```
dt_strong   party_id
 1.397655   1.566525
```

# Example Ordered

## Predicted Probabilities:

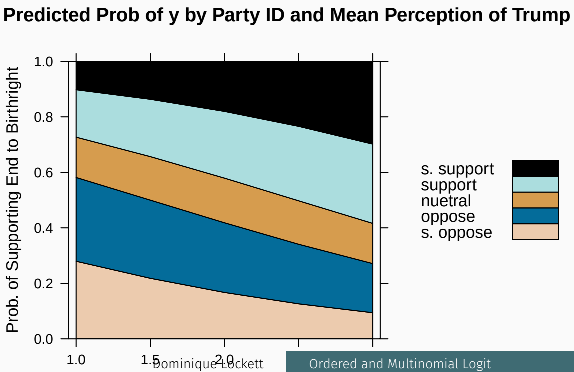
```
> > (Effect(focal.predictors = ('party_id'), m1, given.values = c(dt_strong = mean(dt_strong))))
```

```
party_id effect (probability) for s. oppose
         1        1.5          2        2.5          3
0.27964294 0.21828910 0.16727061 0.12625036 0.09415224

party_id effect (probability) for oppose
        1       1.5         2       2.5         3
0.3019306 0.2816622 0.2510610 0.2147011 0.1770585

party_id effect (probability) for nuetral
        1       1.5         2       2.5         3
0.1452534 0.1568680 0.1609246 0.1566210 0.1448073

party_id effect (probability) for support
        1       1.5         2       2.5         3
0.1710500 0.2066518 0.2405438 0.2683739 0.2858336

party_id effect (probability) for s. support
        1       1.5         2       2.5         3
0.1021230 0.1365289 0.1802000 0.2340536 0.2981484
```

# Example Ordered

## Predicted Probability:

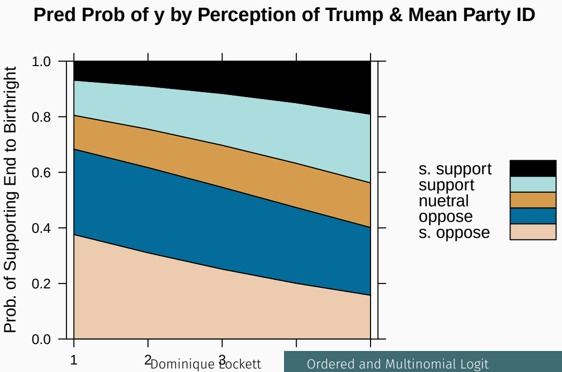
```
#require(wesanderson)
> e.out <— (Effect(focal.predictors = ('party_id'),m1, given.values = c(dt_strong = mean(dt_strong))))
> mean(dt_strong) = 2.508258
>plot(e.out, rug = F, style = 'stacked', main= 'PTitle',key.args = list(space ='right'),ylab = 'Title',
xlab = 'Title',colors = palette(wes_palette("Darjeeling2")))
```



# Example Ordered

## Predicted Probability:

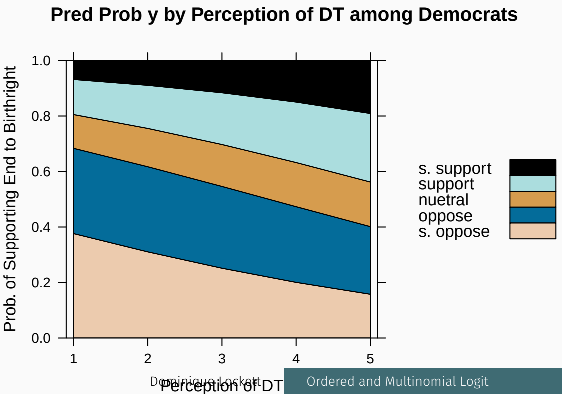
```
#require(wesanderson)
> e.out2 <- (Effect(focal.predictors = ('dt_strong'), m1, given.values = c(party_id = mean(party_id))))
> mean(party_id) = 2.916667
>>plot(e.out2, rug = F, style = 'stacked', main = 'Title', key.args = list(space = 'right'),
xlab = 'Title', ylab = 'Title', colors = palette(wes_palette("Darjeeling2")))
```



# Example Ordered

## Predicted Probability:

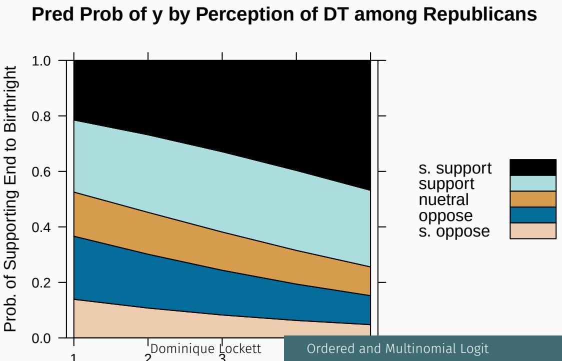
```
#require(wesanderson)
>e.out3 <- (Effect(focal.predictors = ('dt_strong'), m1, given.values = c(party_id = 1)))
>plot(e.out3, rug = F, style = 'stacked',main = 'title', key.args = list(space = 'right'),
xlab = 'title', colors = palette(wes_palette("Darjeeling2")))
```



# Example Ordered

## Predicted Probability:

```
#require(wesanderson)
> e.out4 <— (Effect(focal.predictors = ('dt_strong'), m1, given.values = c(party_id = 3)))
>plot(e.out4, rug = F, style = 'stacked', main = 'Title', key.args = list(space = 'right'),
xlab = 'Title', ylab = 'Title',
colors = palette(wes_palette("Darjeeling2")))
xlab = 'title', colors = palette(wes_palette("Darjeeling2")))
```



# Example Ordered

```
# library(brant)
> brant(model)
```

```
Test for^^IX2^^Idf^^Iprobability

Omnibus^^I^^I12.11^^I6^^I0.06
dt_strong^^I7.36^^I3^^I0.06
party_id^^I6.03^^I3^^I0.11
```

A significant test statistic provides evidence that the parallel regression assumption has been violated.

# Conclusion

# Summary

- Feasible alternatives to linear regression
- Interpreted in log odds
- Useful and basic

# Questions?

V. K. Borooah.
*Logit and probit: ordered and multinominal models.*
Sage, 2001.

W. N. Venables and B. D. Ripley.
*Modern Applied Statistics with S.*
Springer, New York, fourth edition, 2002.
ISBN 0-387-95457-0.

W. N. Venables and B. D. Ripley.
*Modern Applied Statistics with S.*
Springer, New York, fourth edition, 2002.
ISBN 0-387-95457-0.