POL 213 – Lecture 6

Ordered Choice Regression Models[1]

Lauren Peritz

U.C. Davis

May 23, 2024

[1] Thanks to Scott MacKenzie for sharing lecture materials.

Outline

Ordinal response variables

Modeling Polytomous Data

Ordered logit model of tax opinion

Parallel regression assumption

Categorical Dependent Variable Models

Ordinal Response Variables

- In many settings, we want to model categorical response data where the possible outcomes fall into a small number of discrete, ordered categories
- In some cases, the ordered responses arise from grouping continuous data (e.g., income = <$20k, $20k-$39k, $40k-$59k, $60k-$79k, $80k-$99k, $100k+, or time <2/wk, 2-3/wk, 4-6/wk, 7+/wk); in others, the ordered responses reflect assessed data (e.g., Likert scales, support = strongly oppose, somewhat oppose, neutral, somewhat support, strongly support, ideology)
- Ultimately, the ordinal nature of a measure is something the researcher decides for herself
- Some variables can, but probably should not be ordered (e.g., color), others can be ordered in some circumstances, but not others (e.g., partisanship vis-a-vis ideology as opposed to partisanship vis-a-vis state or region)

Ordinal Response Variables

- When the number of response categories is large, using OLS and other models for continuous data can be justified; when the number of categories is small, such models typically work poorly
- Such models assume equal interval scoring, where a unit change in Xi results in a β change in E(Yi); these methods convey a false sense of precision about the relationship between predictors and the response
- The distribution of categories is important as well; OLS will only deliver useful results when the dividing lines between categories are the same distance apart
  - E.g., for income <$20k, $20k-$39k, $40k-$59k, $60k-$79k, $80k-$99k, $100k+ versus <$20k, $20k-$59k, $60k-$74k, $75k-$99k, $100k-$199k, $200k+

Ordinal Response Variables

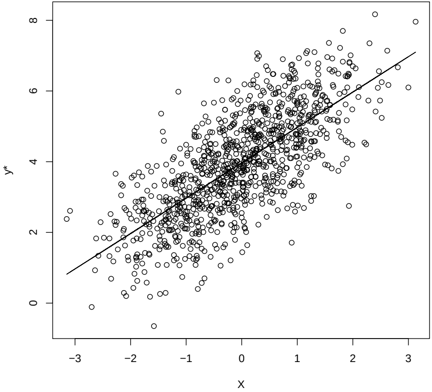
- To see how and why OLS fares poorly, consider the following dataset with X and ε $N(0,1)$ and Y^* generated so that:

$$Y^* = 4 + 1X_i + \varepsilon$$

- If we fit an OLS model to Y^* and plot $\hat{Y}_i^*$ we get:

```
> ols.y <- lm(y ~x)
> summary(ols.y)

Call:
lm(formula = y ~ x)

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.96878    0.03259  121.78   <2e-16
x            1.00043    0.03058   32.72   <2e-16

Residual standard error: 1.029 on 998 degrees of freedom

Multiple R-squared:  0.5175,    Adjusted R-squared:  0.517
F-statistic:  1070 on 1 and 998 DF,  p-value: < 2.2e-16
```

Ordinal Response Variables



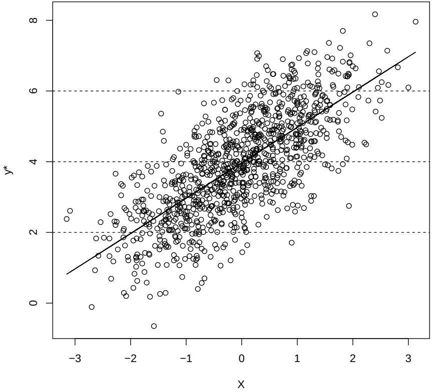
Ordinal Response Variables

- Suppose we did not observe Y^* but a discretized version of it:

$$Y_1 = \begin{cases} 1 & \textit{if} \quad Y^* \leq 2 \\ 2 & \textit{if} \quad 2 < Y^* \leq 4 \\ 3 & \textit{if} \quad 4 < Y^* \leq 6 \\ 4 & \textit{if} \quad Y^* > 6 \end{cases}$$

- The following plot illustrates the partition of Y^* into the four categories of Y_1. The distance between cut points is relatively even although the variability of Y has been cut in half (i.e. from about $\{0, 8\}$ to $\{1, 4\}$).

Ordinal Response Variables



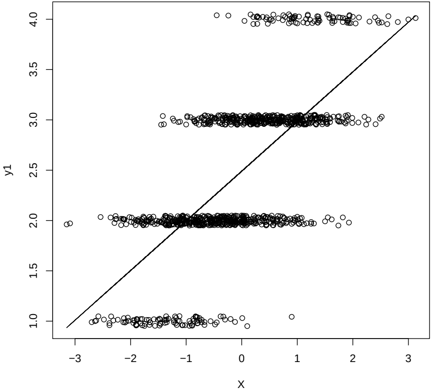
Ordinal Response Variable

If we fit a OLS model to Y1 and plot $\hat{Y1}_i$ we get:

```
> ols.y1 <- lm(y1 ~x)
> summary(ols.y1)
Call:
lm(formula = y1 ~ x)
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.48983    0.01834  135.78   <2e-16
x            0.49345    0.01720   28.68   <2e-16
Residual standard error: 0.5792 on 998 degrees of freedom
Multiple R-squared:  0.4518,    Adjusted R-squared:  0.4513
F-statistic: 822.6 on 1 and 998 DF,  p-value: < 2.2e-16
```

This model is similar to OLS of Y^* albeit with α and β roughly half the size. That reflects the reduced variability in Y.

Ordinal Response Variables



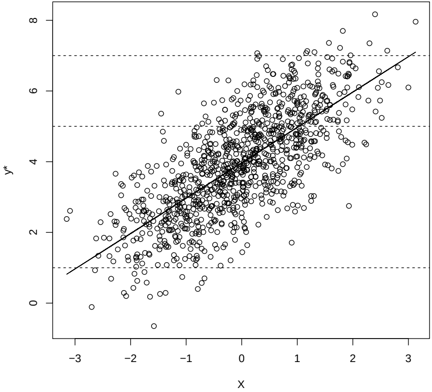
Ordinal Response Variables

Suppose we observe Y2, another discretized version of Y^*

$$Y_2 = \begin{cases} 1 & \textit{if} \quad Y^* \leq 1 \\ 2 & \textit{if} \quad 1 < Y^* \leq 5 \\ 3 & \textit{if} \quad 5 < Y^* \leq 7 \\ 4 & \textit{if} \quad Y^* > 7 \end{cases}$$

The plot on the next slide illustrates this partition of Y^* into the four categories of Y2. The distance between the cut points is uneven, resulting in greater distortion as well as reduced variability of Y.

Ordinal Response Variables



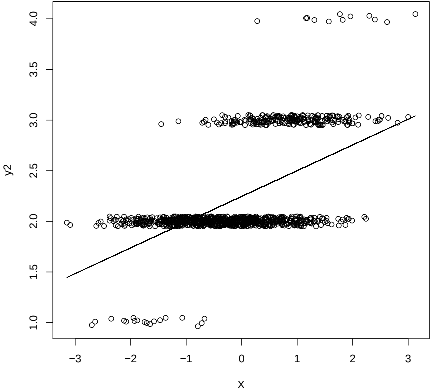
Ordinal Response Variables

- If we fit an OLS model to Y_2 and plot $\hat{Y}_{2,i}$ we get:

```
> ols.y2 <- lm(y2 ~x)
>
> summary(ols.y2)

Call:
lm(formula = y2 ~ x)
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.2463     0.0129   174.2   <2e-16
x              0.2541     0.0121    21.0   <2e-16

Residual standard error: 0.4073 on 998 degrees of freedom
Multiple R-squared:  0.3065,    Adjusted R-squared:  0.3058
F-statistic:   441 on 1 and 998 DF,  p-value: < 2.2e-16
```

- The effect of X is smaller and the lower R^2 suggests this model has a poorer fit than the previous one.

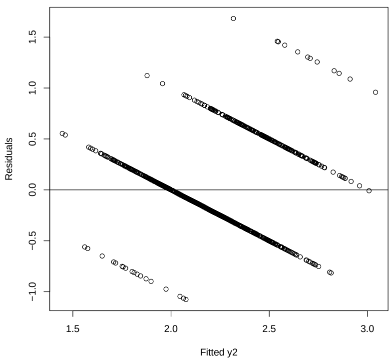
Ordinal Response Variables

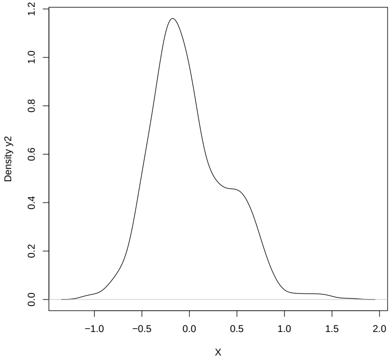


Ordinal Response Variables

- In addition to potentially biased coefficients, applying OLS to ordinal data will result in non-normal heteroskedastic errors
- The next plot shows the fitted values and residuals from the previous model shows the residuals grouped into bands $(1 - \hat{Y_{2,i}}, 2 - \hat{Y_{2,i}}, 3 - \hat{Y_{2,i}}, 4 - \hat{Y_{2,i}})$
- A kernel density plot of the errors suggests they are non-normal, and might be bi-modal
- The bottom line: few categories + asymmetrical cut points means OLS is a poor model choice.

Ordinal Response Variables





Outline

Modeling Polytomous Data II

- We can derive an ordered logit (or probit) model by extending the latent variable formulation for binary regression, which posits an underlying regression for a continuous but unobservable response variable, ξ such that:
  $Y_i = j$ if $\alpha_{j-1} \leq \xi_i < \alpha_j$, where $j \in \{1, ..., J\}$
- For a four-category response variable, we observe:

$$Y_i = \begin{cases} 1 & \textit{if} \quad -\infty \leq \xi_i < \alpha_1 \\ 2 & \textit{if} \quad \alpha_1 \leq \xi_i \leq \alpha_2 \\ 3 & \textit{if} \quad \alpha_2 \leq \xi_i < \alpha_3 \\ 4 & \textit{if} \quad \alpha_3 \leq \xi_i < \infty \end{cases}$$

- We assume this latent variable is a linear function of the explanatory variable and an unobservable error term:

$$\xi_i = \alpha + \beta X_i - \varepsilon_i$$

- We want to estimate α and β but cannot via least squares regression because we do not observe the latent response.

Modeling Polytomous Data II

- Generally, we can write the probability of any particular discrete outcome as:

$$\begin{aligned} P(Y_i = j) &= P(\alpha_{j-1} \leq \xi_i < \alpha_j) \\ &= P(\alpha_{j-1} \leq \alpha + \beta X_i - \varepsilon_i < \alpha_j) \\ &= \int_{-\infty}^{\alpha_j - (\alpha + \beta X_i)} f(\varepsilon_i) d\varepsilon - \int_{-\infty}^{\alpha_{j-1} - (\alpha + \beta X_i)} f(\varepsilon_i) d\varepsilon \\ &= F(\alpha_j - (\alpha + \beta X_i)) - F(\alpha_{j-1} - (\alpha + \beta X_i)) \end{aligned}$$

- Intuitively, we can slice the cumulative density of Y at two points, α_{j-1} and α_j and then calculate the probability that Y_i takes the value associated with this interval by determining the area under the curve between the cut points.
- The Fundamental Theorem of Calculus states that we can do this by integrating up to α_j and then subtracting the area below α_{j-1}.

Modeling polytomous data II

- To make this calculation, we need to choose a distribution for the errors, ε_i. If we assume a standard logistic distribution, $\Lambda(.)$, we get the ordered logit model. If we instead assume a standard normal $\Phi(.)$ we get the ordered probit model.
- Based on this assumption, we can evaluate $P(Y_i = j)\ \forall j$.
- For a four-category response variable, we have:

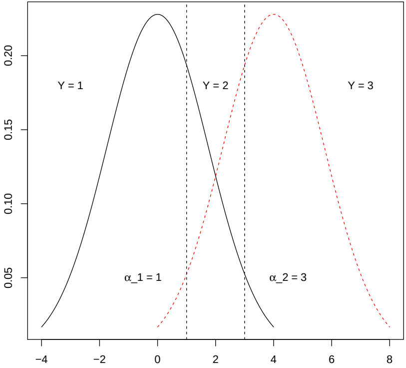
$$\begin{aligned}
P(Y_i = 1) &= \Psi(\alpha_1 - (\alpha + \beta X_i)) - 0 \\
P(Y_i = 2) &= \Psi(\alpha_2 - (\alpha + \beta X_i)) - \Psi(\alpha_1 - (\alpha + \beta X_i)) \\
P(Y_i = 3) &= \Psi(\alpha_3 - (\alpha + \beta X_i)) - \Psi(\alpha_2 - (\alpha + \beta X_i)) \\
P(Y_i = 4) &= 1 - \Psi(\alpha_3 - (\alpha + \beta X_i))
\end{aligned}$$

where ε_i $N(0,1)$, and $\alpha_0 = -\infty$ and $\alpha_1 = \infty$ correspond to probabilities of 0 and 1.

Ordinal response variables

- Note that parameters α and β are invariant to the cut points α_j.
- This is essentially shifting the density of ε along the x-axis while keeping α_j fixed; changing X which changes $\mu = \alpha + \beta X_i$ moves the density of ε relative to the α_j, which changes the probability of various outcomes.
- For example, the following figure illustrates a three-category response variable with the cut points dividing the space into $Y_i = 1$, $Y_i = 2$, and $Y_i = 3$ regions.

Ordinal response variables



Ordinal response variables

- The solid density has most of its mass left of α_1; indeed it is clear that $P(Y_i = 1) > P(Y_i = 2) > P(Y_i = 3)$.
- If a positive change in X increases ξ_i the dashed density illustrates the effect of this change. Now $P(Y_i = 3) > P(Y_i = 2) > P(Y_i = 1)$.
- This is not always true; a smaller shift in the density might have given $P(Y_i = 2) > P(Y_i = 3) > P(Y_i = 1)$.
- While shifting the density right (left) increases (decreases) the probability of the last (first) outcome, the probability of the middle outcomes can increase or decrease when the effect of X on ξ_i is positive.

Modeling Polytomous Data II

- We can estimate this model using MLE by forming the joint probability and writing the log-likelihood; for any single observation the likelihood is:

$$\begin{aligned} L_i &= \textstyle\prod_{j=1}^{J}(P(Y_i = j))^{d_{ij}} \\ &= \textstyle\prod_{j=1}^{J}(F(\alpha_j - (\alpha + \beta X_i)) - F(\alpha_{j-1} - (\alpha + \beta X_i))^{d_{ij}} \end{aligned}$$

  where $d_{ij} = 1$ if $Y_i = j$ and 0 otherwise.
- The likelihood of the data is then:

$$L_i = \prod_{i=1}^{N}\prod_{j=1}^{J}(F(\alpha_j - (\alpha + \beta X_i)) - F(\alpha_{j-1} - (\alpha + \beta X_i))^{d_{ij}}$$

Modeling Polytomous Data II

- The corresponding (generalized) log likelihood is:

$$lnL(\alpha, \beta|y) = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ij} ln(F(\alpha_j - (\alpha + \beta X_i)) - F(\alpha_{j-1} - (\alpha + \beta X_i))$$

- For ordered **probit**, the log likelihood is:

$$lnL(\alpha, \beta|y) = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ij} ln(\Psi(\alpha_j - (\alpha + \beta X_i)) - \Psi(\alpha_{j-1} - (\alpha + \beta X_i))$$

- For ordered **logit**, the log likelihood is:

$$lnL(\alpha, \beta|y) = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ij} ln(\Lambda(\alpha_j - (\alpha + \beta X_i)) - \Lambda(\alpha_{j-1} - (\alpha + \beta X_i))$$

Modeling Polytomous Data II

- In a standard binary regression model, the intercept α is the baseline probability of a success when all X's are 0.
- One might think of the α_j in the ordered choice regression model as a series of additional intercepts; $J - 1$ of these intercepts fully partitions the space.
  - α_j indicates the log-odds of a response equal to or less than J for the baseline group; so α_j increase over J.
- However, estimating a model with all $J - 1$ cut points and an intercept in the model leaves it unidentified.
  - If we included all $J - 1$ cut points in the model, we could shift the intercept up or down (i.e. the location of the density plot). This leaves the probabilities of various outcomes unchanged if we simply shifted the cut points in parallel.
- We can either drop the intercept and estimate all $J - 1$ cut points or keep an intercept term. Stata and R make the choice for you: they both drop the intercept.

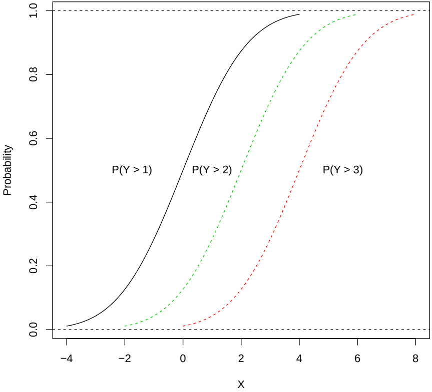
Modeling Polytomous Data II

- Recall the parameters are invariant to the cut points α_j so we only estimate a single β for each predictor.
- Equivalently we assume the effect of X on the probabilities that $Y_i = j$ is the same for all outcomes

$$\frac{\partial P(Y_i = j)}{\partial X} = \frac{\partial P(Y_i = j')}{\partial X} \qquad \forall j \neq j'$$

- This means that the probability curves for the various outcomes will have the same slope; the α_j simply shifts the curve to the right or left (linear transformation).
- The following plut illustrates this parallel slopes assumption for a four-category response variable.

Modeling Polytomous Data II



Modeling Polytomous Data II

- In the ordered logit setting, this assumption is also referred to as the proportional odds assumption.
- To illustrate, let Z be the linear predictor s.t.:

$$Z = \beta_1 X_1 + ... + \beta_k X_k$$

- The probability of the ordered logit model is linear in the log-odds

$$ln \frac{P(Y_i \leq j|X)}{P(Y_i > j|X)} = \alpha_j - Z$$

- Note that the βs are subtracted which means that each β indicates the effect of a one-unit increase in the predictor on the log-odds of a response higher than j.
- For any pair of categories j and j' except the baseline:

$$ln(\frac{\pi_{ij}}{\pi_{ij'}}) = ln \frac{\pi_{ij}/\pi_1}{\pi_{ij'}/\pi_1} = (\alpha_{j'} - \alpha_j) - (Z - Z) = \alpha_{j'} - \alpha_j$$

- For a fixed set of Xs the logits for any pair of categories differ only by the constant $\alpha_{j'} - \alpha_j$. The odds are proportional to one another.

$$\frac{odds_j}{odds_{j'}} = exp(\alpha_{j'} - \alpha_j) = \frac{\exp(\alpha_{j'})}{\exp(\alpha_j)}$$

Modeling Polytomous Data II

- For ordered logit with four-categories, the conditional probabilities of each out j are:

$$
\begin{aligned}
P(Y_i = 1|X) &= \frac{1}{(1 + exp(Z - \alpha_1))} \\
P(Y_i = 2|X) &= \frac{1}{(1 + exp(Z - \alpha_2))} - \frac{1}{(1 + exp(Z - \alpha_1))} \\
P(Y_i = 3|X) &= \frac{1}{(1 + exp(Z - \alpha_3))} - \frac{1}{(1 + exp(Z - \alpha_2))} \\
P(Y_i = 4|X) &= 1 - \frac{1}{(1 + exp(Z - \alpha_3))}
\end{aligned}
$$

- As with binary logit and probit, there is little difference between ordered logit and ordered probit in most practical settings.
- The two models are virtually indistinguishable in probabilities.

Outline

Ordinal response variables

Modeling Polytomous Data

Ordered logit model of tax opinion

Parallel regression assumption

Categorical Dependent Variable Models

Ordered logit model of tax opinion

- Consider example using survey data from 2012 experimental study of CA voters. 795 respondents were asked about their support for nine tax measures considered by the state legislature including a soda tax. Four-category response variable.

```
> w <- table(mydata$tax_soda)
> t = as.data.frame(w)
> names(t)[1] = 'Value'
> t$Opinion <- c("Strongly Oppose", "Somewhat Oppose",
               "Somewhat Support", "Strongly Support")
> t$Percent = t$Freq / 795

> t
Value Freq          Opinion   Percent
1     1  341   Strongly Oppose 0.4289308
2     2   96   Somewhat Oppose 0.1207547
3     3  149  Somewhat Support 0.1874214
4     4  209  Strongly Support 0.2628931
```

- We will look at ideology, income, education, partisanship and party cue[2] on soda tax as predictors.

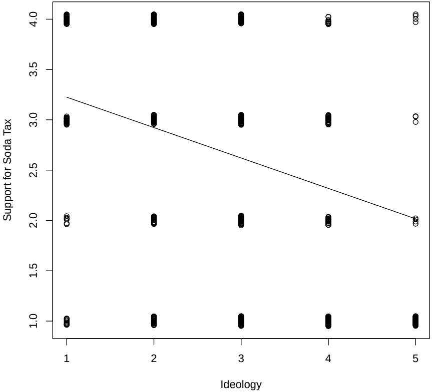
[2] respondents randomly assigned to receive party positions on the tax, DEM support, GOP oppose

Ordered logit model of tax opinion

- The small number of response categories and uneven distribution of responses should signal that OLS is likely to be a poor choice for modeling this data.

```
> ols.soda <- lm(soda_tax ~ ideology + income + education + ind + gop + cue +
                gop_cue, data = ca_soda)

> sum.ols.soda <- summary(ols.soda); sum.ols.soda
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.12071    0.18641  16.741  < 2e-16 ***
ideology    -0.30214    0.04828  -6.258 6.40e-10 ***
income       0.02166    0.01207   1.794  0.07314 .
education    0.09214    0.02867   3.214  0.00136 **
ind         -0.65371    0.12614  -5.183 2.78e-07 ***
gop         -0.85436    0.16934  -5.045 5.63e-07 ***
cue         -0.08939    0.10251  -0.872  0.38348
gop_cue      0.10166    0.17000   0.598  0.55000

Residual standard error: 1.063 on 787 degrees of freedom
Multiple R-squared:  0.2951,    Adjusted R-squared:  0.2888
F-statistic: 47.06 on 7 and 787 DF,  p-value: < 2.2e-16
```

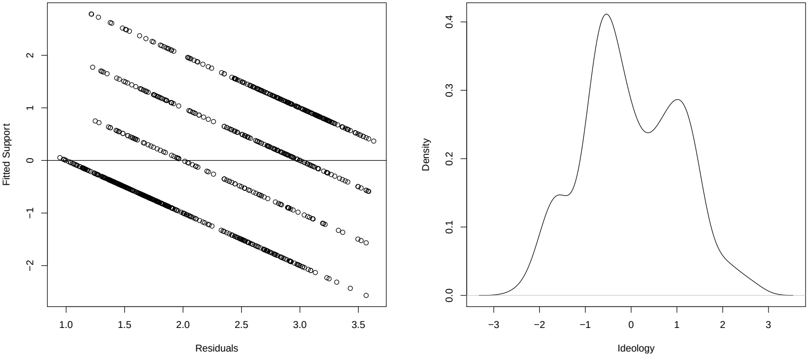
example



Ordered logit model of tax opinion

- The plot of fitted values against the ideology predictor confirms what the low R^2 suggests: this model does not fit the data well.
- A kernel density plot of the errors suggests that they are non-normal and bi-modal.
- Multi-modal error distributions are a common problem when OLS is used on ordinal response data (with few possible outcomes).

example


Fitted Support
Residuals
1.0
1.5
2.0
2.5
3.0
3.5
−2
−1
0
1
2
Density
Ideology
−3
−2
−1
0
1
2
3
0.0
0.1
0.2
0.3
0.4

Ordered logit model of tax opinion

- To estimate the ordered logit model, we can use the `polr()` function in R.
- The output has a single coefficient for each predictor and estimates for the three cut points or intercepts.
- βs are log-odds ratios.
  - The coefficient on ideology (-0.542) indicates that increasing this predictor reduces the log-odds of responding in higher v. lower categories.
- To get the odds ratio, exponentiate the βs:

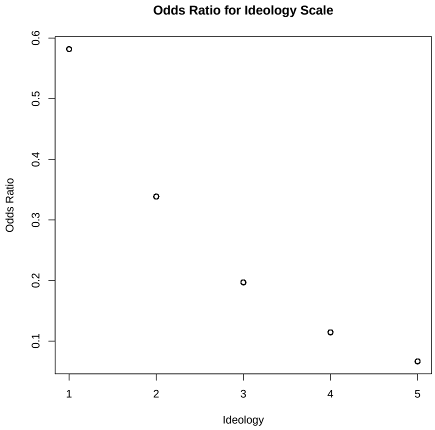
$$\exp(-0.542) = 0.581$$

```
> # Ordered Logit Model of Tax Opinion
>
> ologit.soda <- polr(soda_tax2 ~ ideology + income + education + ind + gop + cue + g
> sum.ologit.soda <- summary(ologit.soda); sum.ologit.soda

Call:
polr(formula = soda_tax2 ~ ideology + income + education + ind +
gop + cue + gop_cue, data = ca_soda)

Coefficients:
Value Std. Error t value
ideology  -0.54158    0.09181 -5.8988
income     0.03936    0.02275  1.7298
education  0.16532    0.05292  3.1239
ind       -1.01203    0.22919 -4.4157
gop       -1.48344    0.32327 -4.5888
cue       -0.16636    0.18307 -0.9087
gop_cue    0.21577    0.33386  0.6463

Intercepts:
Value   Std. Error t value
1|2 -1.8272  0.3514    -5.1990
2|3 -1.1769  0.3477    -3.3843
3|4 -0.1066  0.3458    -0.3084

Residual Deviance: 1774.628
AIC: 1794.628
```

Ordered logit model of tax opinion

- Plotting odds ratios is helpful in assessing how the odds of responding in higher v. lower categories changes as X changes.
- We can use `confint()` to get odds ratios with CIs.
  - The odds of an IND responding in a higher v. lower category are 0.363 times those of DEM (no cue).

```
> or.soda.ci <- exp(cbind(OR = coef(ologit.soda), ci)); or.soda.ci
OR      2.5 %    97.5 %
ideology  0.5818280 0.4852129 0.6956065
income    1.0401415 0.9948077 1.0876766
education 1.1797759 1.0637477 1.3091537
ind       0.3634790 0.2308273 0.5676229
gop       0.2268549 0.1193080 0.4245428
cue       0.8467373 0.5907544 1.2115305
gop_cue   1.2408184 0.6478069 2.4025735
```


Odds Ratio for Ideology Scale
Odds Ratio
0.1
0.2
0.3
0.4
0.5
0.6
1
2
3
4
5
Ideology

Ordered logit model of tax opinion

- We can also calculate odds ratios for relevant profiles.
- The odds ratio for a DEM at the median of ideology, income, and education is 0.410. The odds for a similar DEM in the party cue treatment is 0.347.
- These odds ratios are not that different, i.e. weak treatment.
- On the other hand, the odds ratio for a DEM with liberal views (ideology = 2) is 0.704 and for a DEM with conservative views is 0.238.
- We can flip the interpretation of any odds ratio:
  - For example, the odds of responding in a lower v. higher category are $1/0.23 = 4.196$ for a conservative DEM.

Ordered logit model of tax opinion

```
> x_D_base <- c(3, 6, 3, 0, 0, 0, 0)
> or_D_base <- exp(sum(x_D_base*coef(ologit.soda))); or_D_base
[1] 0.4095784
> x_D_cue <- c(3, 6, 3, 0, 0, 1, 0)
> or_D_cue <- exp(sum(x_D_cue*coef(ologit.soda))); or_D_cue
[1] 0.3468053
>
> x_D_lo <- c(2, 6, 3, 0, 0, 0, 0)
> or_D_lo <- exp(sum(x_D_lo*coef(ologit.soda))); or_D_lo
[1] 0.703951
> x_D_hi <- c(4, 6, 3, 0, 0, 0, 0)
> or_D_hi <- exp(sum(x_D_hi*coef(ologit.soda))); or_D_hi
[1] 0.2383042
```

Ordered Logit Model of Tax Opinion

- We can use the probability formula for ordered logit to calculate the predicted probabilities of support.
- As with binary regression models and mlogit, the probabilities depend on the values X_i and other parameters, thus we should use profiles.
- There are four possible outcomes, each with a probability attached to it.
- We can also use the `predict()` function to get predicted probabilities.

```
> head(yhat.ologit)
1          2          3          4
1 0.2727436 0.14538277 0.25881852 0.32305514
2 0.1745400 0.11378982 0.25325562 0.45841459
3 0.2260003 0.13275794 0.26122538 0.38001642
4 0.5942191 0.14302860 0.15384476 0.10890759
5 0.6608270 0.12789834 0.12714346 0.08413120
6 0.8239452 0.07572684 0.06349457 0.03683334
```

Ordered Logit Model of Tax Opinion

```
> # Probabilities
> zeta.vector <- cbind(-1.8271604, -1.1768830, -0.1066465); zeta.vector
[,1]      [,2]       [,3]
[1,] -1.82716 -1.176883 -0.1066465
>
> z_D_base <- sum(x_D_base * coef(ologit.soda))
> p1_D_base <- 1 / (1 + exp(z_D_base - zeta.vector[1])); p1_D_base
[1] 0.2820059
> p2_D_base <- 1 / (1 + exp(z_D_base - zeta.vector[2])) - 1 / (1 + exp(z_D_base - zet
[1] 0.1474048
> p3_D_base <- 1 / (1 + exp(z_D_base - zeta.vector[3])) - 1 / (1 + exp(z_D_base - zet
[1] 0.2575569
> p4_D_base <- 1 - 1 / (1 + exp(z_D_base - zeta.vector[3])); p4_D_base
[1] 0.3130324
>
> p1_D_base + p2_D_base + p3_D_base + p4_D_base
[1] 1
```

Ordered logit model of tax opinion

- For a DEM at the median of the predictors, the predicted probability of "strongly opposing" is 0.28, "somewhat opposing" is 0.15, "somewhat supporting" is 0.26, and "strongly supporting" is 0.31.
- For a DEM with relatively **liberal** views, the predicted probabilities are $(0.19, 0.11, 0.26, 0.44)$. For a DEM with relatively **conservative** views, the predicted probabilities are $(0.40, 0.16, 0.23, 0.21)$.
- Those predicted probabilities sum to 1.

Ordered logit model of tax opinion

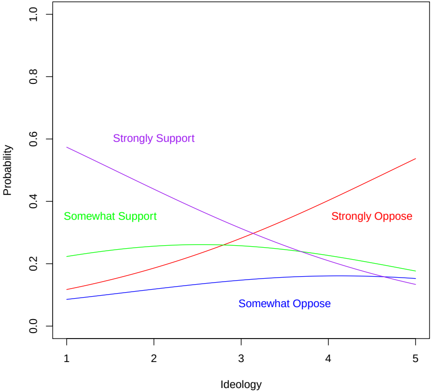
```
> z_D_lo <- sum(x_D_lo * coef(ologit.soda))

> p1_D_lo <- 1 / (1 + exp(z_D_lo - zeta.vector[1])); p1_D_lo
[1] 0.1860151
> p2_D_lo <- 1 / (1 + exp(z_D_lo - zeta.vector[2])) - 1 / (1 + exp(z_D_lo - zeta.vect
[1] 0.118511
> p3_D_lo <- 1 / (1 + exp(z_D_lo - zeta.vector[3])) - 1 / (1 + exp(z_D_lo - zeta.vect
[1] 0.2562715
> p4_D_lo <- 1 - 1 / (1 + exp(z_D_lo - zeta.vector[3])); p4_D_lo
[1] 0.4392023

> z_D_hi <- sum(x_D_hi * coef(ologit.soda))

> p1_D_hi <- 1 / (1 + exp(z_D_hi - zeta.vector[1])); p1_D_hi
[1] 0.4030066
> p2_D_hi <- 1 / (1 + exp(z_D_hi - zeta.vector[2])) - 1 / (1 + exp(z_D_hi - zeta.vect
[1] 0.1609718
> p3_D_hi <- 1 / (1 + exp(z_D_hi - zeta.vector[3])) - 1 / (1 + exp(z_D_hi - zeta.vect
[1] 0.2264584
> p4_D_hi <- 1 - 1 / (1 + exp(z_D_hi - zeta.vector[3])); p4_D_hi
[1] 0.2095631
```

Ordered logit model of tax opinion

- Having calculated predicted probabilities, we can plot the probabilities of giving each response against ideology.
- Support for the tax generally decreases in ideology.

```
> plot(x, base_soda[,1], type="l", ylim=c(0,1), xlab="Ideology",
+      ylab="Probability", col = "red")
> lines(x, base_soda[,2], lty = 1, col = "blue")
> lines(x, base_soda[,3], lty = 1, col = "green")
> lines(x, base_soda[,4], lty = 1, col = "purple")
> text(4.5, 0.35, "Strongly Oppose", col="red")
> text(3.5, 0.07, "Somewhat Oppose", col="blue")
> text(1.5, 0.35, "Somewhat Support", col="green")
> text(2.0, 0.60, "Strongly Support", col="purple")
```

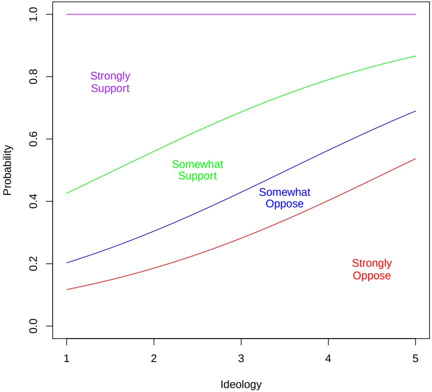
Ordered logit model of tax opinion



Ordered logit model of tax opinion

- When the response variable taxes on several possible values, a graphical representation of the fitted model can aid interpretation.
  - The following plot shows cumulative probabilities of support for different values of ideology.
  - A clear majority of liberal DEMS support the soda tax.
- Similar displays can also help us interpret interaction terms in our model.

Ordered logit model of tax opinion



Ordered logit model of tax opinion

- As with binary regression and mlogit models, we can use simulations from the multivariate normal to randomly draw parameter values and then calculate the mean probabilities with uncertainty estimates.
- Or use `Zelig` to calculate predicted values, probabilities, and first differences

```
> z.out <- zelig(soda_tax2 ~ ideology + income + education + ind + gop + cue + gop_cu
> x.out1a <- setx(z.out, ideology = 3, income = 6, education = 3, ind = 0, gop = 0, c
> x.out1b <- setx(z.out, ideology = 3, income = 6, education = 5, ind = 0, gop = 0, c
> s.out1ab <- sim(z.out, x = x.out1a, x1 = x.out1b)
> summary(s.out1ab)
fd
mean          sd          50%         2.5%        97.5%
1 -0.06245336 0.019053050 -0.06180224 -0.09978130 -0.024970303
2 -0.01605881 0.007789744 -0.01506715 -0.03246938 -0.003860434
3  0.01046356 0.021676076  0.00364246 -0.01526306  0.068717541
4  0.06804862 0.022348148  0.06853297  0.02350267  0.112862354
```

Ordered logit model of tax opinion

- The same formula we used for binary logit can be used to calculate Wald statistics (z-scores) and confidence intervals for the multinomial logit.

```
> z <- summary(ologit.soda)$coefficients[,1] / summary(ologit.soda)$coefficients[,2]
> z

ideology      income   education        ind         gop         cue     gop_cue
-5.8988358  1.7298494  3.1239244 -4.4156796 -4.5888143 -0.9087427  0.6462880
1|2         2|3          3|4
-5.1990168 -3.3843285 -0.3084212

> p <- (1 - pnorm(abs(z), 0, 1))*2; p
ideology        income    education          ind           gop           cue
3.660753e-09 8.365719e-02 1.784563e-03 1.006931e-05 4.457708e-06 3.634859e-01
gop_cue          1|2          2|3          3|4
5.180929e-01 2.003454e-07 7.135259e-04 7.577619e-01
```

Ordered logit model of tax opinion

- For the interaction terms, we can use the `glht()` function to test the joint significance of the multiple predictors.

```
> summary(glht(ologit.soda, linfct = c("cue + gop_cue = 0")))

Simultaneous Tests for General Linear Hypotheses
Fit: polr(formula = soda_tax2 ~ ideology + income + education + ind +
gop + cue + gop_cue, data = ca_soda, Hess = TRUE)
Linear Hypotheses:
Estimate Std. Error z value Pr(>|z|)
cue + gop_cue == 0  0.04941    0.27769   0.178    0.859
(Adjusted p values reported -- single-step method)
```

Ordered logit model of tax opinion

Calculate first differences by comparing predicted effects at various levels

Variable	**Comparison**	**Str. Opp**	**M. Opp.**	**M. Sup.**	**Str. Sup.**
Ideology	(LQ,UQ)	**0.214**	**0.038**	-0.032	**-0.221**
Income	(LQ,UQ)	-0.039	-0.007	0.004	0.036
Education	(LQ,UQ)	**-0.060**	**-0.015**	0.002	**0.068**
IND	(0,1) at X_{med}	**0.239**	0.006	**-0.072**	**-0.162**
GOP	(0,1) at X_{med}	**0.353**	-0.012	**-0.111**	**-0.213**
Cue	(0,1) at X_{med}	0.033	0.005	-0.002	-0.029
GOP*Cue	(0,1) at GOP=1	-0.048	0.008	0.016	0.013

Ordered logit model of tax opinion

- We can compute the pseudo-R^2 as we did for binary logit and mlogit.

```
> # Pseudo-R2
>
> pR2(ologit.soda)
llh         llhNull            G2       McFadden          r2ML          r2CU
-887.3138545 -1020.2968743   265.9660398     0.1303376     0.2843388     0.30798
```

- We can also adjust for differences in the degrees of freedom by estimating the null model and computing an adjusted McFadden statistic.

```
> # Adjusted McFadden
>
> L.base <- logLik(ologit.soda2)
>
> L.full <- logLik(ologit.soda)
> P3 <- attr(L.full, "df")
>
> McFadden.R2.full <- 1 - (L.full / L.base); McFadden.R2.full
'log Lik.' 0.1303376 (df=10)
> McFadden.Adj.R2.full <- 1 - ((L.full - P3) / L.base); McFadden.Adj.R2.full
'log Lik.' 0.1205365 (df=10)
```

- The pseudo-R^2 values suggest the model is superior to the null[3]

[3] Some skepticism due.

Ordered logit model of tax opinion

- We can also calculate PRE statistics for the ologit model
- The model with ideology, income, education, partisanship, and party cues reduces error by approximately 20.4%.
- The simulated PRE is slightly lower, 19.8% with confidence interval (18.1%, 21.4%)
- Both the AIC and BIC measures indicate that the model is superior to the null, admittedly not a high bar.

Ordered logit model of tax opinion

```
> pre(ologit.soda, sim=TRUE, R=1000)
mod1:  soda_tax2 ~ ideology + income + education + ind + gop + cue + gop_cue
mod2:  soda_tax2 ~ 1

Analytical Results
PMC =  0.429
PCP =  0.546
PRE =  0.205
ePMC =  0.303
ePCP =  0.406
ePRE =  0.148

Simulated Results
median lower upper
PRE 0.198  0.181 0.214
ePRE 0.147  0.126 0.169
>
> # Information measures
> AIC(ologit.soda2)
[1] 2046.594
> AIC(ologit.soda)
[1] 1794.628
```

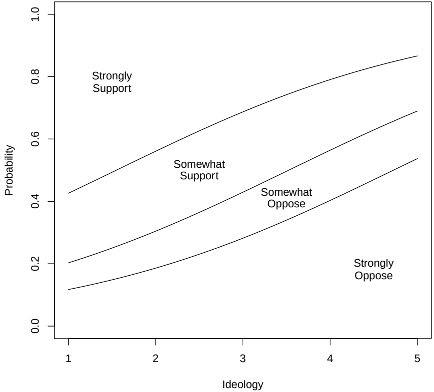
Outline

Parallel regression assumption

- Both the ordered logit and ordered probit models assume that the relationship between the predictors and the response are the same for each pair of outcome categories (e.g. the lowest versus all higher categories, the next lowest versus all higher categories, etc.)
- Because the relationships are the same, we estimate a single β for each predictor with the α_j summarizing the cumulative distribution of responses (for the baseline where all Xs are 0).
- This is a restrictive assumption that is useful when it holds; when it does all we need to know about how a predictor moves respondents across the scale of the response is β.
- When the assumption does not hold, it is possible that one or more predictors has a differential effect over the range of the response, e.g. stronger above versus below some value.

Parallel regression assumption

- How might we examine the validity of this assumption?
- We could separate our response into J-1 binary response variables (e.g. $P(Y > 1)$, $P(Y > 2)$, $P(Y > 3)$) and estimate a binary logit for each. We could then compare the coefficients for the J-1 models and see if they are similar across the different partitions of the ordinal response variable.
- Tests of the parallel regression assumption adopt this approach.



Parallel regression assumption

- Let's examine whether the parallel regression assumption holds for the soda tax.
- We can estimate the soda tax model using `ologit` and then use the 'brant' package in R .
- This test is an automated version of the Wald Test (Brant 1990; Long 1997). The output includes estimates for the J-1 binary logits.
- The test calculates predictor-specific Wald tests.

```
>
> brant(ologit.soda)
--------------------------------------------
Test for X2 df probability
--------------------------------------------
Omnibus 17.09 14 0.25
ideology 3.57 2 0.17
income 5.53 2 0.06
education 1.62 2 0.45
ind 0.63 2 0.73
gop 2.57 2 0.28
cue 3.96 2 0.14
gop_cue 1.03 2 0.6
--------------------------------------------

H0: Parallel Regression Assumption holds
```

Parallel regression assumption

- The Wald test suggests that the parallel regression assumption is reasonable for the model. This is often not the case.
- The output (following slide) is from an ologit model of support for a change of ownership tax; the proposal was designed to prevent companies from avoiding property tax reassessment by parceling ownership out to supporters.
- As with the soda tax, the effects of ideology, education and partisanship are large and significant.

```
> ca_property <- subset(d, select = c("tax_owner_change", "ideology", "income", "educ
> ologit.property <- polr(as.factor(tax_owner_change) ~ ideology + income + educate +
> summary(ologit.property)
Call:
polr(formula = as.factor(tax_owner_change) ~ ideology + income +
    educate + ind + gop + cue + gop_cue, data = ca_property,
    Hess = TRUE)

Coefficients:
           Value Std. Error t value
ideology -0.70703    0.09495  -7.446
income    0.01177    0.02272   0.518
educate   0.18063    0.05369   3.365
ind      -0.87709    0.24352  -3.602
gop      -1.75465    0.32809  -5.348
cue      -0.26878    0.19786  -1.358
gop_cue   0.71847    0.32652   2.200

Intercepts:
    Value   Std. Error t value
1|2 -3.5003  0.3759    -9.3124
2|3 -2.4267  0.3637    -6.6720
3|4 -0.7330  0.3543    -2.0688

Residual Deviance: 1643.696
AIC: 1663.696
(286 observations deleted due to missingness)
```

Parallel regression assumption

- The Brant test suggests that the parallel regression assumption is less appropriate for this model.
  - Coefficients for ideology, IND, GOP, education, etc. vary across the binary logit models.
- The global Wald statistic is significant as are the statistics for ideology, income and IND.
- The simple ordered logit model may not be a good fit for these data.

Parallel regression assumption

```
> brant(ologit.property)
--------------------------------------------
Test for X2 df probability
--------------------------------------------
Omnibus 38.47 14 0
ideology 6.23 2 0.04
income 5.46 2 0.07
educate 4.21 2 0.12
ind 7.21 2 0.03
gop 1.62 2 0.45
cue 3.09 2 0.21
gop_cue 2.92 2 0.23
--------------------------------------------

H0: Parallel Regression Assumption holds
```

Parallel regression assumption

- What to do if a model violates the parallel regression assumption:
  - Nothing (a common practice)
  - Use a non-ordinal model like multinomial logit
  - Estimate an alternative ordinal model that relaxes the assumption
- Several alternatives exist for modeling ordinal response data when the parallel regression assumption is violated.
  - Generalized Ordered Logit relaxes the assumption by estimating a separate vector of coefficients for each of the J-1 response categories.
    (+) relaxes modeling assumptions but (-) introduces many more parameters.
  - Partial Proportional Odds models relax the assumption for some predictors (e.g. those with violations) while retaining it for others.
- Proceed with caution about your research conclusions.

Outline

Categorical Dependent Variables

Stuff we've covered or at least mentioned...

1. Binary (logit; probit)
2. Ordinal (ordinal logit; ordinal probit)
3. Unordered multi-category (multinomial logit; multinomial probit)

Extended topics in Maximum Likelihood Estimation (MLE)...

4. Count data (Poisson, Zero-inflated Poisson, Negative Binomial)
5. Duration/survival models (Cox Proportional Hazard)
6. Censored data (Tobit)
7. Selection models (Heckman)