ORDINAL LOGISTIC REGRESSION

Dr. Aric LaBarr
Institute for Advanced Analytics

INTRODUCTION

Logistic Regression

- What if there are more than two categories?
  - Ordinal Logistic Regression
  - Multinomial Logistic Regression

- When the outcomes are **ordered** we can generalize the binary logistic regression model.
- Examples:
  - Disagree, Neutral, Agree
  - Tropical Depression, Tropical Storm, Category 1, 2, 3, 4, 5 Hurricanes

Ordinal Logistic Regression

- Models are used when the response variable is ordinal.

- Models can also be used when the continuous response variable has a **restricted range** and need to be split into categories.

Logistic Models

- Binary Logistic Regression (probability that observation *i* has the event):

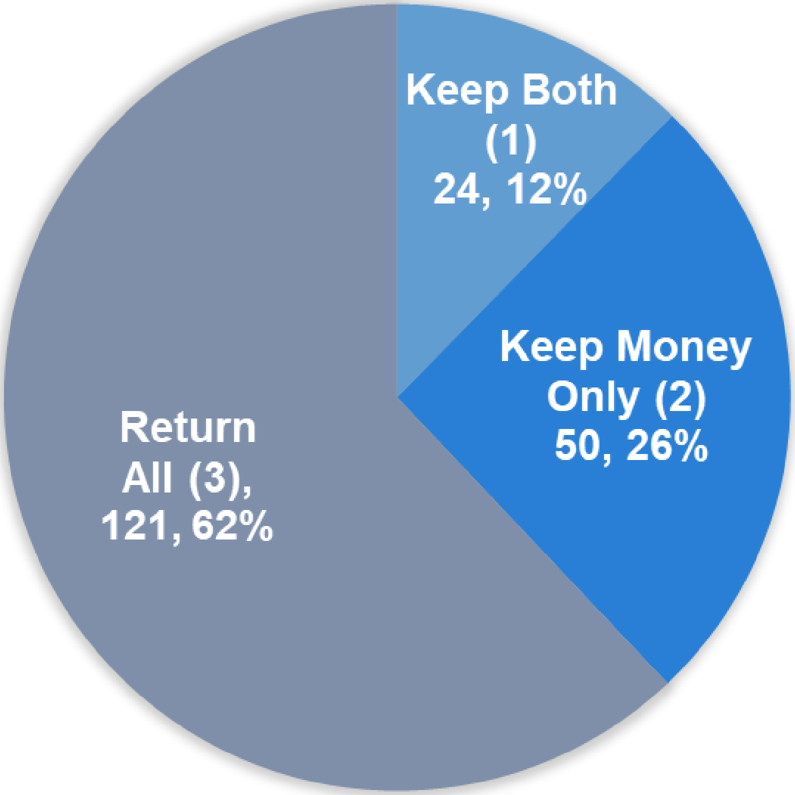
$$= \beta_0 + \beta_1 x_{1,i} + \cdots \beta_k x_{k,i}$$

- Ordinal Logistic Regression (probability that observation *i* has **at most** event *j*, and $j = 1, \ldots, m$):

$$= \beta_{0,j} + \beta_1 x_{1,i} + \cdots \beta_k x_{k,i}$$

"Found a Wallet?" Data Set

- Model the association between various factors and different levels of ethical responses on finding a wallet.
- 195 observations in the data set.



"Found a Wallet?" Data Set

- Model the association between various factors and different levels of ethical responses on finding a wallet.
- Students at Upenn.
- Predictors:
  - **male:** indicator for a male student
  - **business:** indicator for student enrolled in business school
  - **punish:** how often the student was punished as a child – low (1), moderate (2), high (3)
  - **explain:** indicator of whether explanation for punishment was given

PROPORTIONAL ODDS MODEL

Methods for Modeling

- There are three methods for modeling ordinal logistic regression models:
  1. Cumulative Logit Model
  2. Adjacent Categories Model
  3. Continuation Ratio Model

Methods for Modeling

- There are three methods for modeling ordinal logistic regression models:
  1. Cumulative Logit Model
  2. Adjacent Categories Model
  3. Continuation Ratio Model

Easy to implement and interpret! Also, most common…

Cumulative Logits

- Instead of modeling the typical logit, we will model the cumulative logits.
- If an ordinal variable has *m* levels with probabilities $(p_1, p_2, \ldots, p_m)$, then the cumulative logits are:

$$\log\left(\frac{p_{i,1}}{p_{i,2} + p_{i,3} + \cdots p_{i,m}}\right), \log\left(\frac{p_{i,1} + p_{i,2}}{p_{i,3} + \cdots + p_{i,m}}\right), \ldots, \log\left(\frac{p_{i,1} + \cdots + p_{i,m-1}}{p_m}\right)$$

Cumulative Logits

- Instead of modeling the typical logit, we will model the cumulative logits.
- If an ordinal variable has *m* levels with probabilities $(p_1, p_2, \ldots, p_m)$, then the cumulative logits are:

$$\boxed{\log\left(\frac{p_{i,1}}{p_{i,2} + p_{i,3} + \cdots p_{i,m}}\right)}, \log\left(\frac{p_{i,1} + p_{i,2}}{p_{i,3} + \cdots + p_{i,m}}\right), \ldots, \log\left(\frac{p_{i,1} + \cdots + p_{i,m-1}}{p_m}\right)$$

$$p_{i,1} + p_{i,2} + \cdots + p_{i,m} = 1$$

Cumulative Logits

- Instead of modeling the typical logit, we will model the cumulative logits.
- If an ordinal variable has *m* levels with probabilities $(p_1, p_2, \ldots, p_m)$, then the cumulative logits are:

$$\boxed{\log\left(\frac{p_{i,1}}{p_{i,2} + p_{i,3} + \cdots p_{i,m}}\right)}, \log\left(\frac{p_{i,1} + p_{i,2}}{p_{i,3} + \cdots + p_{i,m}}\right), \ldots, \log\left(\frac{p_{i,1} + \cdots + p_{i,m-1}}{p_m}\right)$$

$$p_{i,1} + p_{i,2} + \cdots + p_{i,m} = 1 \longrightarrow p_{i,2} + \cdots + p_{i,m} = 1 - p_{i,1}$$

Cumulative Logits

- Instead of modeling the typical logit, we will model the cumulative logits.
- If an ordinal variable has *m* levels with probabilities $(p_1, p_2, \ldots, p_m)$, then the cumulative logits are:

$$\boxed{\log\left(\frac{p_{i,1}}{1-p_{i,1}}\right)}, \log\left(\frac{p_{i,1}+p_{i,2}}{p_{i,3}+\cdots+p_{i,m}}\right), \ldots, \log\left(\frac{p_{i,1}+\cdots+p_{i,m-1}}{p_m}\right)$$

$$p_{i,1}+p_{i,2}+\cdots+p_{i,m}=1 \longrightarrow p_{i,2}+\cdots+p_{i,m}=1-p_{i,1}$$

Cumulative Logits

- Instead of modeling the typical logit, we will model the cumulative logits.
- If an ordinal variable has *m* levels with probabilities $(p_1, p_2, \ldots, p_m)$, then the cumulative logits are:

$$\log\left(\frac{p_{i,1}}{1-p_{i,1}}\right), \boxed{\log\left(\frac{p_{i,1}+p_{i,2}}{p_{i,3}+\cdots+p_{i,m}}\right)}, \ldots, \log\left(\frac{p_{i,1}+\cdots+p_{i,m-1}}{p_m}\right)$$

Cumulative Logits

- Instead of modeling the typical logit, we will model the cumulative logits.
- If an ordinal variable has *m* levels with probabilities $(p_1, p_2, \dots, p_m)$, then the cumulative logits are:

$$\log\left(\frac{p_{i,1}}{1-p_{i,1}}\right), \boxed{\log\left(\frac{p_{i,1}+p_{i,2}}{p_{i,3}+\cdots+p_{i,m}}\right)}, \dots, \log\left(\frac{p_{i,1}+\cdots+p_{i,m-1}}{p_m}\right)$$

$$p_{i,1}+p_{i,2}+\cdots+p_{i,m}=1 \longrightarrow p_{i,3}+\cdots+p_{i,m}=1-\left(p_{i,1}-p_{i,2}\right)$$

Cumulative Logits

- Instead of modeling the typical logit, we will model the cumulative logits.
- If an ordinal variable has *m* levels with probabilities $(p_1, p_2, \dots, p_m)$, then the cumulative logits are:

$$\log\left(\frac{p_{i,1}}{1-p_{i,1}}\right), \boxed{\log\left(\frac{p_{i,1}+p_{i,2}}{1-\left(p_{i,1}+p_{i,2}\right)}\right)}, \dots, \log\left(\frac{p_{i,1}+\cdots+p_{i,m-1}}{p_m}\right)$$

$$p_{i,1}+p_{i,2}+\cdots+p_{i,m}=1 \longrightarrow p_{i,3}+\cdots+p_{i,m}=1-\left(p_{i,1}-p_{i,2}\right)$$

Cumulative Logits

- Instead of modeling the typical logit, we will model the cumulative logits.
- If an ordinal variable has *m* levels with probabilities $(p_1, p_2, \ldots, p_m)$, then the cumulative logits are:

$$\log\left(\frac{p_{i,1}}{1-p_{i,1}}\right), \log\left(\frac{p_{i,1}+p_{i,2}}{1-(p_{i,1}+p_{i,2})}\right), \ldots, \boxed{\log\left(\frac{p_{i,1}+\cdots+p_{i,m-1}}{1-(p_{i,1}+\cdots+p_{i,m-1})}\right)}$$

Cumulative Logits

- Instead of modeling the typical logit, we will model the cumulative logits.
- If an ordinal variable has *m* levels with probabilities $(p_1, p_2, \dots, p_m)$, then the cumulative logits are:

$$\log\left(\frac{p_{i,1}}{1-p_{i,1}}\right), \log\left(\frac{p_{i,1}+p_{i,2}}{1-(p_{i,1}+p_{i,2})}\right), \dots, \log\left(\frac{p_{i,1}+\cdots+p_{i,m-1}}{1-(p_{i,1}+\cdots+p_{i,m-1})}\right)$$

***m-1* Binary Logistic Regressions!**

- Event now becomes outcome $\leq j$ for categories $j = 1, \dots, m$

Logistic Models

- Binary Logistic Regression (probability that observation *i* has the event):

$$= \beta_0 + \beta_1 x_{1,i} + \cdots \beta_k x_{k,i}$$

- Ordinal Logistic Regression (probability that observation *i* has **at most** event m, and $j = 1, \ldots, m$):

$$= \beta_{0,j} + \beta_1 x_{1,i} + \cdots \beta_k x_{k,i}$$

m – 1 Equations!

Logistic Models

- Binary Logistic Regression (probability that observation *i* has the event):

$$= \beta_0 + \beta_1 x_{1,i} + \cdots \beta_k x_{k,i}$$

- Ordinal Logistic Regression (probability that observation *i* has **at most** event m, and $j = 1, \dots, m$):

$$= \beta_{0,j} + \beta_1 x_{1,i} + \cdots \beta_k x_{k,i}$$

- Intercept changes, but **slope parameters stays the same**!

"Found a Wallet?" Data Set

- Model the association between various factors and different levels of ethical responses on finding a wallet.

$$\log\left(\frac{p_{i,1}}{p_{i,2}+p_{i,3}}\right) = \beta_{0,1} + \beta_1 \mathrm{male}_i + \beta_2 \mathrm{business}_i + \beta_3 \mathrm{punishM}_i + \beta_4 \mathrm{punishH}_i + \beta_5 \mathrm{explain}_i$$

$$\log\left(\frac{p_{i,1}+p_{i,2}}{p_{i,3}}\right) = \beta_{0,2} + \beta_1 \mathrm{male}_i + \beta_2 \mathrm{business}_i + \beta_3 \mathrm{punishM}_i + \beta_4 \mathrm{punishH}_i + \beta_5 \mathrm{explain}_i$$

Proportional Odds Model – SAS

```
proc logistic data=Logistic.Wallet;
   class punish(param=ref ref='1');
   model wallet = male business punish explain
                       / clodds=pl;
run;
```

Proportional Odds Model – SAS

The LOGISTIC Procedure

Model Information	
Data Set	LOGISTIC.WALLET
Response Variable	wallet
Number of Response Levels	3
Model	cumulative logit
Optimization Technique	Fisher's scoring

Number of Observations Read	195
Number of Observations Used	195

Proportional Odds Model – SAS

Response Profile		
Ordered Value	**wallet**	**Total Frequency**
1	1	24
2	2	50
3	3	121

Probabilities modeled are cumulated over the lower Ordered Values.

Class Level Information			
Class	**Value**	**Design Variables**	
punish	**1**	0	0
	2	1	0
	3	0	1

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Proportional Odds Model – SAS

Score Test for the Proportional Odds Assumption		
Chi-Square	**DF**	**Pr > ChiSq**
5.2215	5	0.3894

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	356.140	321.335
SC	362.686	344.246
-2 Log L	352.140	307.335

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	44.8047	5	<.0001
Score	40.9509	5	<.0001
Wald	38.5978	5	<.0001

Proportional Odds Model – SAS

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
male	1	10.6047	0.0011
business	1	4.4167	0.0356
punish	2	9.4185	0.0090
explain	1	9.4925	0.0021

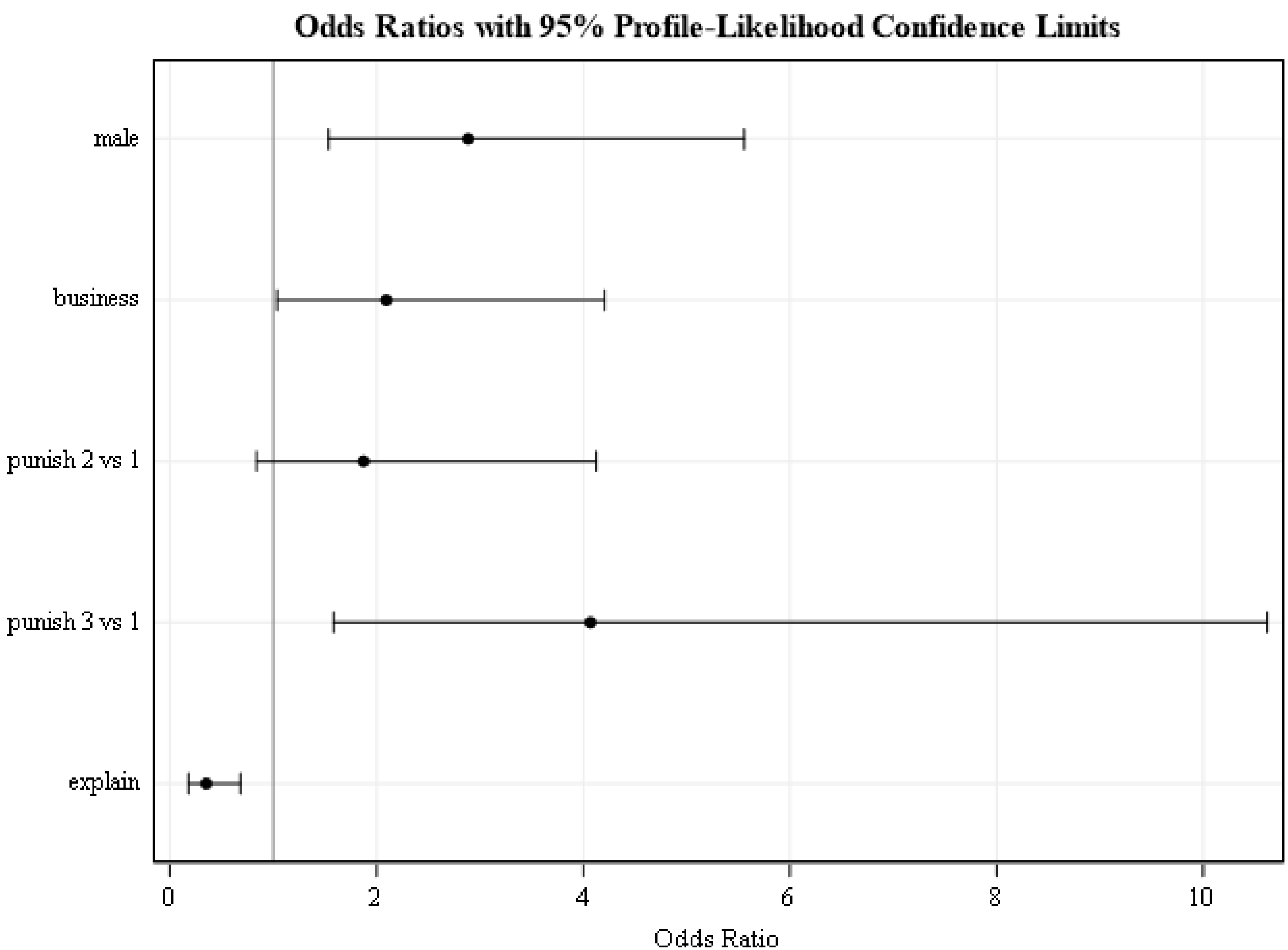
Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	**1**	1	-2.5678	0.4169	37.9321	<.0001
Intercept	**2**	1	-0.7890	0.3675	4.6107	0.0318
male		1	1.0598	0.3254	10.6047	0.0011
business		1	0.7389	0.3516	4.4167	0.0356
punish	**2**	1	0.6277	0.4005	2.4564	0.1170
punish	**3**	1	1.4031	0.4721	8.8330	0.0030
explain		1	-1.0518	0.3414	9.4925	0.0021

Proportional Odds Model – SAS

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	69.0	Somers' D	0.465
Percent Discordant	22.5	Gamma	0.508
Percent Tied	8.5	Tau-a	0.249
Pairs	10154	c	0.732

Odds Ratio Estimates and Profile-Likelihood Confidence Intervals				
Effect	Unit	Estimate	95% Confidence Limits	
male	1.0000	2.886	1.533	5.557
business	1.0000	2.094	1.039	4.206
punish 2 vs 1	1.0000	1.873	0.838	4.124
punish 3 vs 1	1.0000	4.068	1.583	10.616
explain	1.0000	0.349	0.178	0.681

Proportional Odds Model – SAS



Proportional Odds Model – R

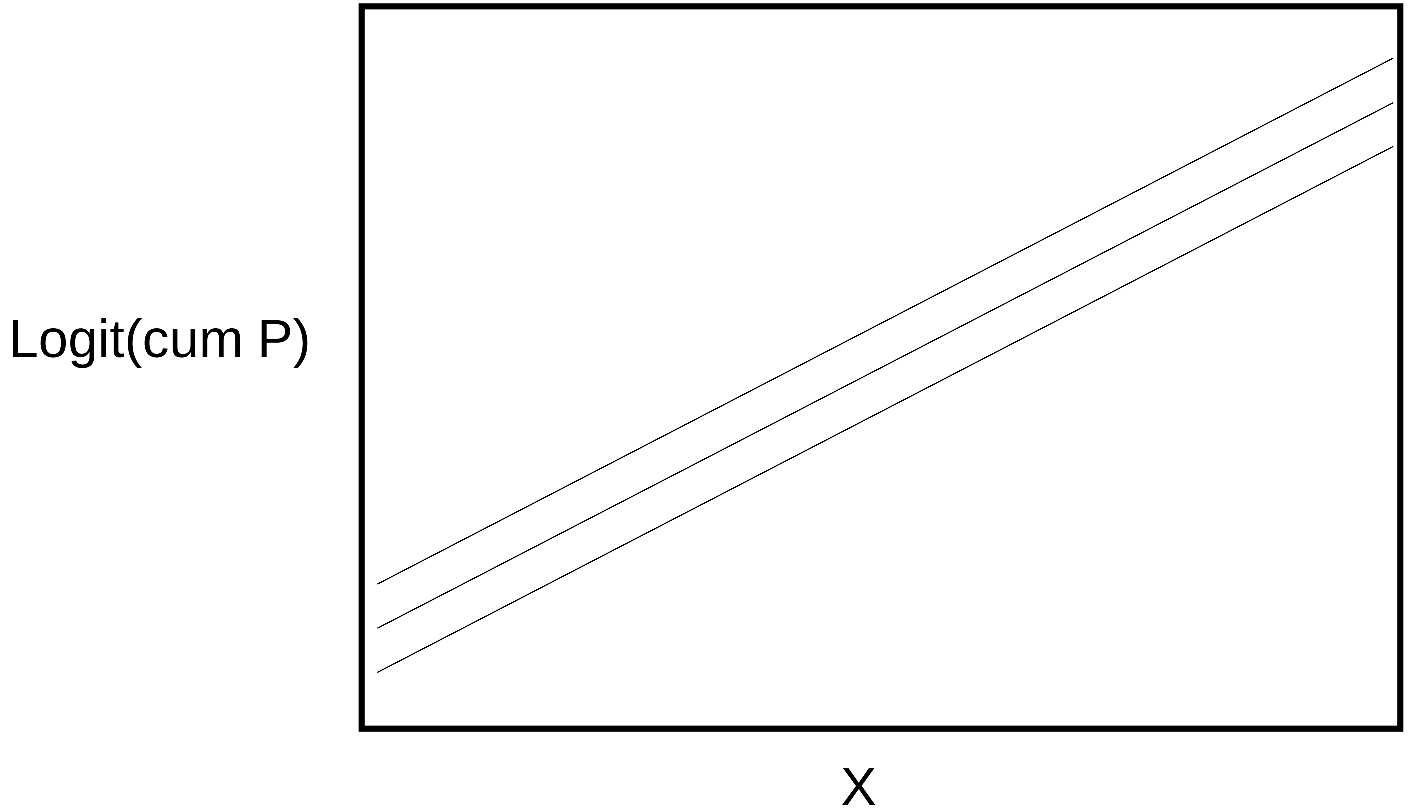
```
train <- wallet
train$punish <- factor(train$punish)

clogit.model <- polr(factor(wallet) ~ male + business + punish +
                     explain,
                     method = "logistic", data = train)
summary(clogit.model)
```

Proportional Odds Model – R

```
## Coefficients:
##             Value Std. Error t value
## male     -1.0598     0.3274  -3.237
## business -0.7389     0.3556  -2.078
## punish2  -0.6276     0.4048  -1.551
## punish3  -1.4031     0.4823  -2.909
## explain   1.0519     0.3408   3.086
##
## Intercepts:
##     Value   Std. Error t value
## 1|2 -2.5679  0.4190    -6.1287
## 2|3 -0.7890  0.3709    -2.1273
##
## Residual Deviance: 307.3349
## AIC: 321.3349
```

Testing Assumptions



HOW DO WE TEST IF SLOPES ARE THE SAME?

Score Test for Proportional Odds

- Need to test to see if the slopes are statistically different from each other in the proportional odds model.
  - Null: Proportional Odds Correct (Slopes Equal Across Models)
  - Alternative: Proportional Odds Incorrect (Slopes NOT Equal Across Models)

Testing Assumption – SAS Default

Score Test for the Proportional Odds Assumption		
Chi-Square	**DF**	**Pr > ChiSq**
5.2215	5	0.3894

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	356.140	321.335
SC	362.686	344.246
-2 Log L	352.140	307.335

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	44.8047	5	<.0001
Score	40.9509	5	<.0001
Wald	38.5978	5	<.0001

Testing Assumption – R

```
brant(clogit.model)

## --------------------------------------------
## Test for X2   df   probability
## --------------------------------------------
## Omnibus       5.46    5   0.36
## male      0.51    1   0.47
## business 0.58    1   0.45
## punish2       0.99    1   0.32
## punish3       2.81    1   0.09
## explain       0.25    1   0.62
## --------------------------------------------
##
## H0: Parallel Regression Assumption holds

##                  X2 df probability
## Omnibus  5.4618058  5  0.36215220
## male     0.5123944  1  0.47410417
## business 0.5791753  1  0.44663576
## punish2  0.9871507  1  0.32043977
## punish3  2.8104051  1  0.09365472
## explain  0.2468865  1  0.61927599
```

Testing Assumption – SAS Option

```
proc logistic data=Logistic.Wallet;
   class punish(param=ref ref='1');
   model wallet = male business punish explain /
                  unequalslopes clodds=pl;
   male: test male_1 = male_2;
   business: test business_1 = business_2;
   punish2: test punish2_1 = punish2_2;
   punish3: test punish3_1 = punish3_2;
   explain: test explain_1 = explain_2;
   title 'Ordinal Logistic Regression - Unequal Slopes';
run;
quit;
```

Testing Assumption – SAS Option

Linear Hypotheses Testing Results			
Label	**Wald Chi-Square**	**DF**	**Pr > ChiSq**
male	0.7582	1	0.3839
business	0.9922	1	0.3192
punish2	0.7561	1	0.3845
punish3	3.0150	1	0.0825
explain	0.2850	1	0.5935

What if Assumption Fails?

- The proportional odds assumption may not be met for all variables.
- 2 Approaches:
  1. Partial Proportional Odds Model
  2. Multinomial Logistic Regression

INTERPRETATION

Model Notation

- With cumulative logits, increasing the right-hand side of the equation leads to an increased log(odds) of **lower** outcome category:

$$\log\left(\frac{p_{i,1}}{p_{i,2}+p_{i,3}}\right)=\beta_{0,1}+\beta_1\text{male}_i+\beta_2\text{business}_i$$
$$+\beta_3\text{punishM}_i+\beta_4\text{punishH}_i+\beta_5\text{explain}_i$$

$$\log\left(\frac{p_{i,1}+p_{i,2}}{p_{i,3}}\right)=\beta_{0,2}+\beta_1\text{male}_i+\beta_2\text{business}_i$$
$$+\beta_3\text{punishM}_i+\beta_4\text{punishH}_i+\beta_5\text{explain}_i$$

Model Notation

- With cumulative logits, increasing the right-hand side of the equation leads to an increased log(odds) of **lower** outcome category:

$$\log\left(\frac{\boxed{p_{i,1}}}{p_{i,2}+p_{i,3}}\right) = \beta_{0,1} + \beta_1 \text{male}_i + \beta_2 \text{business}_i + \beta_3 \text{punishM}_i + \beta_4 \text{punishH}_i + \beta_5 \text{explain}_i$$

$$\log\left(\frac{\boxed{p_{i,1}+p_{i,2}}}{p_{i,3}}\right) = \beta_{0,2} + \beta_1 \text{male}_i + \beta_2 \text{business}_i + \beta_3 \text{punishM}_i + \beta_4 \text{punishH}_i + \beta_5 \text{explain}_i$$

Model Notation – SAS Default

- With cumulative logits, increasing the right-hand side of the equation leads to an increased log(odds) of **lower** outcome category:

$$\log\left(\frac{\boxed{p_{i,1}}}{p_{i,2}+p_{i,3}}\right) = \beta_{0,1} + \beta_1 \text{male}_i + \beta_2 \text{business}_i + \beta_3 \text{punishM}_i + \beta_4 \text{punishH}_i + \beta_5 \text{explain}_i$$

$$\log\left(\frac{\boxed{p_{i,1}+p_{i,2}}}{p_{i,3}}\right) = \beta_{0,2} + \beta_1 \text{male}_i + \beta_2 \text{business}_i + \beta_3 \text{punishM}_i + \beta_4 \text{punishH}_i + \beta_5 \text{explain}_i$$

Odds Ratio Interpretation – Ascending

- Interpretation is still an odds ratio: $\mathbf{100 * \left(e^{\hat{\beta}_j} - 1\right)\%}$ **higher expected odds** of being in a lower category.

No Association

Group in denominator has higher odds to be in a **lower** category.

Group in numerator has higher odds to be in a **lower** category.

0

1

∞

Odds Ratio Interpretation – Ascending

- Interpretation is still an odds ratio: $\mathbf{100 * \left(e^{\hat{\beta}_j} - 1\right)\%}$ **higher expected odds** of being in a lower category.
- Proportional odds model:
  - Same increase in odds across all singular jumps in category.
  - Wallet example: OR same comparing 1 to 2,3 and from 1,2 to 3.

Proportional Odds Model – SAS

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	69.0	Somers' D	0.465
Percent Discordant	22.5	Gamma	0.508
Percent Tied	8.5	Tau-a	0.249
Pairs	10154	c	0.732

Odds Ratio Estimates and Profile-Likelihood Confidence Intervals				
Effect	Unit	Estimate	95% Confidence Limits	
male	1.0000	2.886	1.533	5.557
business	1.0000	2.094	1.039	4.206
punish 2 vs 1	1.0000	1.873	0.838	4.124
punish 3 vs 1	1.0000	4.068	1.583	10.616
explain	1.0000	0.349	0.178	0.681

Model Notation – SAS Option

- With cumulative logits, increasing the right-hand side of the equation leads to an increased log(odds) of **higher** outcome category:

$$\log\left(\frac{p_{i,1}}{p_{i,2}+p_{i,3}}\right) = \beta_{0,1} - \beta_1 \mathrm{male}_i - \beta_2 \mathrm{business}_i$$
$$-\beta_3 \mathrm{punishM}_i - \beta_4 \mathrm{punishH}_i - \beta_5 \mathrm{explain}_i$$

$$\log\left(\frac{p_{i,1}+p_{i,2}}{p_{i,3}}\right) = \beta_{0,2} - \beta_1 \mathrm{male}_i - \beta_2 \mathrm{business}_i$$
$$-\beta_3 \mathrm{punishM}_i - \beta_4 \mathrm{punishH}_i - \beta_5 \mathrm{explain}_i$$

Odds Ratio Interpretation – Descending

- Interpretation is still an odds ratio: $\mathbf{100 * \left(e^{\widehat{\beta}_j} - 1\right)\%}$ **higher expected odds** of being in a higher category.

No Association

Group in denominator has higher odds to be in a **higher** category.

Group in numerator has higher odds to be in a **higher** category.

0

1

∞

PREDICTIONS AND DIAGNOSTICS

Similarities

- Ordinal logistic regression has a lot of the same aspects/issues as a binary logistic regression:
  - Multicollinearity still exists.
  - Non-convergence problems still exist.
  - Confidence intervals need profile likelihoods.
  - Concordance, Discordance, Tied pairs still exist – so the c statistic still exists.
  - Generalized R^2 remains the same.

Differences

- Ordinal logistic regression has a few aspects/issues that differ from a binary logistic regression:
  - A lot of the diagnostics for binary regression cannot be calculated easily since there are actually **multiple** models – ROC curves for each model?
  - Diagnostics / Influence plots are not available – residuals for each model?
  - Predicted probabilities are for **each** category.

Predicted Probabilities – SAS

```
proc logistic data=Logistic.Wallet;
   class punish(param=ref ref='1');
   model wallet = male business punish explain;
   output out=pred predprobs=I;
   title 'Ordinal Logistic Regression - Ascending';
run;

proc print data=pred;
run;

proc freq data=pred;
   tables _from_*_into_ / plots=none;
   title 'Cross Table of Observed by Predicted Responses';
run;
```

Predicted Probabilities – SAS

Obs	wallet	male	business	punish	explain	_FROM_	_INTO_	IP_1	IP_2	IP_3
1	2	0	0	2	0	2	3	0.12563	0.33412	0.54026
2	2	0	0	2	1	2	3	0.04779	0.18135	0.77087
3	3	0	0	1	1	3	3	0.02609	0.11086	0.86305
4	3	0	0	2	0	3	3	0.12563	0.33412	0.54026
5	1	1	0	1	1	1	3	0.07177	0.24233	0.68591
6	3	0	0	1	1	3	3	0.02609	0.11086	0.86305
7	3	0	0	1	1	3	3	0.02609	0.11086	0.86305
8	3	1	0	1	1	3	3	0.07177	0.24233	0.68591
9	3	1	0	1	1	3	3	0.07177	0.24233	0.68591
10	3	0	0	2	1	3	3	0.04779	0.18135	0.77087

Predicted Probabilities – R

```
pred_probs <- predict(clogit.model, newdata = train, type = "prob
s")
print(pred_probs)
```

```
##                1          2          3
## 1   0.12562481 0.3341195 0.54025570
## 2   0.04778463 0.1813420 0.77087342
## 3   0.02609095 0.1108549 0.86305415
## 4   0.12562481 0.3341195 0.54025570
## 5   0.07176375 0.2423258 0.68591049
## 6   0.02609095 0.1108549 0.86305415
## 7   0.02609095 0.1108549 0.86305415
## 8   0.07176375 0.2423258 0.68591049
## 9   0.07176375 0.2423258 0.68591049
## 10  0.04778463 0.1813420 0.77087342
```

Confusion Matrix

- A confusion matrix is a matrix of all predicted responses compared to actual responses in terms of correct percentage.

	Predicted		
Actual	4	11	9
	3	9	38
	0	12	109

Confusion Matrix

- A confusion matrix is a matrix of all predicted responses compared to actual responses in terms of correct percentage.

	Predicted		
Actual	16.7%	45.8%	37.5%
	6.0%	18.0%	76.0%
	0.0%	9.9%	90.1%

Good Confusion Matrix

	Predicted		
Actual	100%	0.0%	0.0%
	0.0%	100%	0.0%
	0.0%	0.0%	100%

Confusion Matrix

- Weighted accuracy scores are common.

	Predicted		
Actual	1	0.5	0
	0.5	1	0.5
	0	0.5	1