CHAPTER 3: Ordered and unordered multinomial Logit models

Prof. Alan Wan
1. Introduction

2. Unordered multinomial Logit model
   2.1 Categorical and Multinomial distributions
   2.2 Multinomial Logit model: basic idea
   2.3 PROC CATMOD
   2.4 Multinomial Logit model for contingency tables
      2.4.1 Contingency tables
      2.4.2 Class exercise
   2.5 General form

3. Ordered multinomial Logit model
   3.1 Basic idea and latent variable
   3.2 Estimation
Introduction

Outcomes are often characterised by more than two discrete choices; for example

► Which major, accounting, finance, economics, management sciences, will a university student choose, given his/her grades and interests?

► Which country, Hong Kong, Singapore, Korea, Japan, will a firm locate its Asian head office in, given the characteristics of the firm?

► How would you rate the food at City U’s student canteen, excellent, good, fair, poor, very poor?

► How likely do you think China will surpass the U.S. as the world’s No.1 economy in ten years? Very likely, not sure, not likely?
Introduction

- Responses in the last two of the above examples are "ordinal" in that the categories can be ordered in a meaningful way, whereas in the first two examples, the responses have no apparent ordering;
- Unordered multinomial (or simply multinomial) Logit model is appropriate for responses with no ordering;
- Ordered multinomial (or simply ordered or cumulative) Logit model is appropriate for responses with ordering;
- Both multinomial and ordered Logits are generalisations of the binary Logit, but neither the multinomial Logit nor the ordered Logit generalises the other.
Categorical and Multinomial distributions

- The multinomial Logit model assumes that the response follows a Categorical distribution;
Categorical and Multinomial distributions

- The multinomial Logit model assumes that the response follows a Categorical distribution;
- Basic properties of the Categorical distribution:
  1. \( s > 0 \) number of categories, and the events \( Y = 1, Y = 2, \ldots, Y = s \) are mutually exclusive;
  2. \( Pr(Y = y) = p_y, y = 1, 2, \ldots, s \) with \( \sum_{y=1}^{s} p_y = 1 \);
  3. The probability density function of \( Y \) is
     \[
     Pr(Y = y) = p_1 I(y=1) p_2 I(y=2) \ldots p_s I(y=s),
     \]
     where \( I(A) \) is an indicator function taking on 1 if \( A \) is true, 0 otherwise.
Categorical and Multinomial distributions

- The multinomial Logit model assumes that the response follows a Categorical distribution;
- Basic properties of the Categorical distribution:
  1. $s > 0$ number of categories, and the events $Y = 1, Y = 2, ..., Y = s$ are mutually exclusive;
  2. $Pr(Y = y) = p_y, y = 1, 2, ..., s$ with $\sum_{y=1}^{s} p_y = 1$;
  3. The probability density function of $Y$ is $Pr(Y = y) = p_1^{I(y=1)} p_2^{I(y=2)} .... p_s^{I(y=s)}$, where $I(A)$ is an indicator function taking on 1 if $A$ is true, 0 otherwise.
- Clearly, the Bernoulli distribution is a special case of the Categorical distribution with $s=2$ resulting in $y=1,2$ (or $y=0,1$ based on (0,1) indexing)
2.1 Categorical and Multinomial distributions

Let $W_y$ be a random variable representing the number of times in $n$ trials for which the outcome falls in class $y$. Note that $\sum_{y=1}^{s} w_y = n$;
Categorical and Multinomial distributions

Let $W_y$ be a random variable representing the number of times in $n$ trials for which the outcome falls in class $y$. Note that $\sum_{y=1}^{s} w_y = n$;

$W_y$ follows a Multinomial distribution. The joint probability density function of $W_1, W_2, \ldots, W_s$ is given by

$$Pr(W_1 = w_1, W_2 = w_2, \ldots, W_s = w_s) = \frac{n!}{w_1!w_2!\ldots w_s!} p_1^{w_1} p_2^{w_2} \ldots p_s^{w_s}$$
Categorical and Multinomial distributions

Let $W_y$ be a random variable representing the number of times in $n$ trials for which the outcome falls in class $y$. Note that $\sum_{y=1}^{s} w_y = n$;

$W_y$ follows a Multinomial distribution. The joint probability density function of $W_1, W_2, ...., W_s$ is given by

$$Pr(W_1 = w_1, W_2 = w_2, ...., W_s = w_s) = \frac{n!}{w_1!w_2!....w_s!} p_1^{w_1} p_2^{w_2} .... p_s^{w_s}$$

The Binomial distribution is a special case of the Multinomial distribution with $s = 2$;
Categorical and Multinomial distributions

- Let $W_y$ be a random variable representing the number of times in $n$ trials for which the outcome falls in class $y$. Note that $\sum_{y=1}^{s} w_y = n$;  
- $W_y$ follows a Multinomial distribution. The joint probability density function of $W_1$, $W_2$, .... $W_s$ is given by

$$Pr(W_1 = w_1, W_2 = w_2, ..... W_s = w_s) = \frac{n!}{w_1!w_2!....w_s!} p_1^{w_1} p_2^{w_2} ..... p_s^{w_s}$$

- The Binomial distribution is a special case of the Multinomial distribution with $s = 2$;
- The Categorical distribution is a special case of the Multinomial distribution with $n = 1$
Categorical and Multinomial distributions

Consider an example. A box contains 5 red balls, 4 white balls and 3 blue balls. A ball is selected at random from the box, its colour is noted and then the ball is replaced. Find the probability that out of 6 balls selected, 3 are red, 2 are white and 1 is blue.

\[ Pr(red) = \frac{5}{12}, \quad Pr(white) = \frac{4}{12}, \quad Pr(blue) = \frac{3}{12} \]

Hence \[ Pr(3\ red,\ 2\ white,\ 1\ blue) = \frac{6!}{3!2!1!}(\frac{5}{12})^3(\frac{4}{12})^2(\frac{3}{12})^1 \]
\[ = \frac{625}{5184} \]
Example: Survey of 195 undergraduates at the University of Pennsylvania in order to study the effects of parenting styles on altruistic behaviour.

- Question of interest: if you found a wallet on the street, would you
  1. Keep the wallet and the money ($WALLET = 1$)?
  2. Keep the money and return the wallet ($WALLET = 2$)?
  3. Return both the wallet and the money ($WALLET = 3$)?
Multinomial Logit model: basic idea

Example: Survey of 195 undergraduates at the University of Pennsylvania in order to study the effects of parenting styles on altruistic behaviour.

- Question of interest: if you found a wallet on the street, would you
  1. Keep the wallet and the money \((WALLET = 1)\)?
  2. Keep the money and return the wallet \((WALLET = 2)\)?
  3. Return both the wallet and the money \((WALLET = 3)\)?

- The frequencies of the responses are 24, 50 and 121 for
  \(WALLET = 1, 2\) and 3 respectively;
Multinomial Logit model: basic idea

Example: Survey of 195 undergraduates at the University of Pennsylvania in order to study the effects of parenting styles on altruistic behaviour.

▶ Question of interest: if you found a wallet on the street, would you

1. Keep the wallet and the money (\(\text{WALLET} = 1\))?
2. Keep the money and return the wallet (\(\text{WALLET} = 2\))?
3. Return both the wallet and the money (\(\text{WALLET} = 3\))?

▶ The frequencies of the responses are 24, 50 and 121 for \(\text{WALLET} = 1, 2\) and 3 respectively;

▶ Note that \(\text{WALLET}\) is an ordinal variable but let us ignore its ordering feature for the time being.
Possible explanatory variables are:

1. **MALE:** 1 = male; 0 = female;
2. **BUSINESS:** 1 = enrolled in business school, 0 = otherwise;
3. **PUNISH:** A variable describing whether the student was physically punished by parents at various ages:
   - 1 = punished in elementary school but not in middle or high school;
   - 2 = punished in elementary school and middle school but not in high school;
   - 3 = punished at all three levels
4. **EXPLAIN:** When you were punished, did your parents explain to you why you were wrong?
   - 1 = almost always, 0 = sometimes or never
Multinomial Logit model: basic idea

- Define $p_{ij} = Pr(WALLET_i = j)$, $i = 1, \cdots, n$, $j = 1, 2, 3$. 
Multinomial Logit model: basic idea

- Define $p_{ij} = Pr(WALLET_i = j)$, $i = 1, \cdots, n, j = 1, 2, 3$.
- Define the logs of the odds:

  $$
  \ln\left(\frac{p_{i1}}{p_{i3}}\right) = Z_{i1} = \beta_{11} + \beta_{21} MALE_i + \beta_{31} BUSINESS_i + \beta_{41} PUNISH_i + \beta_{51} EXPLAIN_i
  $$

  $$
  \ln\left(\frac{p_{i2}}{p_{i3}}\right) = Z_{i2} = \beta_{12} + \beta_{22} MALE_i + \beta_{32} BUSINESS_i + \beta_{42} PUNISH_i + \beta_{52} EXPLAIN_i
  $$

  $$
  \ln\left(\frac{p_{i1}}{p_{i2}}\right) = Z_{i3} = \beta_{13} + \beta_{23} MALE_i + \beta_{33} BUSINESS_i + \beta_{43} PUNISH_i + \beta_{53} EXPLAIN_i
  $$
Multinomial Logit model: basic idea

- One of the three equations is redundant; for example the third equation may be obtained by

\[
\ln\left( \frac{p_{i1}}{p_{i2}} \right) = \ln\left( \frac{p_{i1}}{p_{i3}} \right) - \ln\left( \frac{p_{i2}}{p_{i3}} \right);
\]
Multinomial Logit model: basic idea

One of the three equations is redundant; for example the third equation may be obtained by
\[ \ln\left(\frac{p_{i1}}{p_{i2}}\right) = \ln\left(\frac{p_{i1}}{p_{i3}}\right) - \ln\left(\frac{p_{i2}}{p_{i3}}\right); \]

Hence
\[ \beta_{i3} = \beta_{11} - \beta_{12}, \]
\[ \beta_{23} \text{MALE}_i = (\beta_{21} - \beta_{22}) \text{MALE}_i; \]
\[ \beta_{33} \text{BUSINESS}_i = (\beta_{31} - \beta_{32}) \text{BUSINESS}_i; \]
\[ \beta_{43} \text{PUNISH}_i = (\beta_{41} - \beta_{42}) \text{PUNISH}_i; \]
\[ \beta_{53} \text{EXPLAIN}_i = (\beta_{51} - \beta_{52}) \text{EXPLAIN}_i; \]
Multinomial Logit model: basic idea

- Solving for these three probabilities based on the two log of the odds equations, we obtain

\[ p_{i1} = \frac{e^{Z_{i1}}}{1 + e^{Z_{i1}} + e^{Z_{i2}}} \]
\[ p_{i2} = \frac{e^{Z_{i2}}}{1 + e^{Z_{i1}} + e^{Z_{i2}}} \]
\[ p_{i3} = \frac{1}{1 + e^{Z_{i1}} + e^{Z_{i2}}} \]
Multinomial Logit model: basic idea

- Solving for these three probabilities based on the two log of the odds equations, we obtain

\[ p_{i1} = \frac{e^{Z_{i1}}}{1 + e^{Z_{i1}} + e^{Z_{i2}}} \]
\[ p_{i2} = \frac{e^{Z_{i2}}}{1 + e^{Z_{i1}} + e^{Z_{i2}}} \]
\[ p_{i3} = \frac{1}{1 + e^{Z_{i1}} + e^{Z_{i2}}} \]

- One can immediately verify that
  - the sum of the three probabilities is 1, and
Multinomial Logit model: basic idea

- Solving for these three probabilities based on the two log of the odds equations, we obtain

\[
p_{i1} = \frac{e^{Z_{i1}}}{1 + e^{Z_{i1}} + e^{Z_{i2}}} \\
p_{i2} = \frac{e^{Z_{i2}}}{1 + e^{Z_{i1}} + e^{Z_{i2}}} \\
p_{i3} = \frac{1}{1 + e^{Z_{i1}} + e^{Z_{i2}}}
\]

- One can immediately verify that
  - the sum of the three probabilities is 1, and
  - when \( j = 2 \), the probability expressions reduce to those under the binary Logit model.
PROC CATMOD

DATA WALLET;
INFILE 'D:\TEACHING\MS4225\WALLET.TXT';
input WALLET MALE BUSINESS PUNISH EXPLAIN;
PROC CATMOD DATA=WALLET;
DIRECT MALE BUSINESS PUNISH EXPLAIN;
MODEL WALLET=MALE BUSINESS PUNISH EXPLAIN/NOITER COVB;
RUN;
The CATMOD Procedure

Data Summary

Response WALLET Response Levels 3
Weight Variable None Populations 23
Data Set WALLET Total Frequency 195
Frequency Missing 0 Observations 195

Population Profiles

Sample MALE BUSINESS PUNISH EXPLAIN Sample Size
1 0 0 1 0 12
2 0 0 1 1 50
3 0 0 2 0 7
4 0 0 2 1 7
5 0 0 3 0 5
6 0 0 3 1 3
7 0 1 1 0 1
8 0 1 1 1 5
9 0 1 2 0 1
10 0 1 2 1 3
11 0 1 3 0 3
12 1 0 1 0 9
13 1 0 1 1 40
14 1 0 2 0 8
15 1 0 2 1 3
16 1 0 3 0 3
17 1 0 3 1 3
18 1 1 1 0 5
19 1 1 1 1 16
20 1 1 2 0 2
21 1 1 2 1 4
22 1 1 3 0 2
23 1 1 3 1 3

Response Profiles

Response WALLET
1 1
2 2
3 3
2. Unordered multinomial Logit model

2.1 Categorical and Multinomial distributions

2.2 Multinomial Logit model: basic idea

2.3 PROC CATMOD

2.4 Multinomial Logit model for contingency tables

2.5 General form

PROC CATMOD

Maximum Likelihood Analysis

Maximum likelihood computations converged.

The CATMOD Procedure

Maximum Likelihood Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2</td>
<td>17.79</td>
<td>0.0001</td>
</tr>
<tr>
<td>MALE</td>
<td>2</td>
<td>12.28</td>
<td>0.0022</td>
</tr>
<tr>
<td>BUSINESS</td>
<td>2</td>
<td>4.69</td>
<td>0.0960</td>
</tr>
<tr>
<td>PUNISH</td>
<td>2</td>
<td>10.92</td>
<td>0.0043</td>
</tr>
<tr>
<td>EXPLAIN</td>
<td>2</td>
<td>9.77</td>
<td>0.0076</td>
</tr>
</tbody>
</table>

Likelihood Ratio 36 31.95 0.6619

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Function Number</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1 1</td>
<td>-3.4712</td>
<td>0.8439</td>
<td>16.92</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>2 2</td>
<td>-1.2917</td>
<td>0.6073</td>
<td>4.52</td>
<td>0.0334</td>
</tr>
<tr>
<td>MALE</td>
<td>1 1</td>
<td>1.2673</td>
<td>0.5546</td>
<td>5.22</td>
<td>0.0223</td>
</tr>
<tr>
<td></td>
<td>2 2</td>
<td>1.1699</td>
<td>0.3715</td>
<td>9.92</td>
<td>0.0016</td>
</tr>
<tr>
<td>BUSINESS</td>
<td>1 1</td>
<td>1.1804</td>
<td>0.5486</td>
<td>4.63</td>
<td>0.0314</td>
</tr>
<tr>
<td></td>
<td>2 2</td>
<td>0.4179</td>
<td>0.4233</td>
<td>0.97</td>
<td>0.3235</td>
</tr>
<tr>
<td>PUNISH</td>
<td>1 1</td>
<td>1.0817</td>
<td>0.3335</td>
<td>10.52</td>
<td>0.0012</td>
</tr>
<tr>
<td></td>
<td>2 2</td>
<td>0.1957</td>
<td>0.2889</td>
<td>0.46</td>
<td>0.4981</td>
</tr>
<tr>
<td>EXPLAIN</td>
<td>1 1</td>
<td>-1.6006</td>
<td>0.5455</td>
<td>8.61</td>
<td>0.0033</td>
</tr>
<tr>
<td></td>
<td>2 2</td>
<td>-0.8040</td>
<td>0.4034</td>
<td>3.97</td>
<td>0.0463</td>
</tr>
</tbody>
</table>
## Introduction

2. Unordered multinomial Logit model

3. Ordered multinomial Logit model

### 2.2 Multinomial Logit model: basic idea

#### 2.4 Multinomial Logit model for contingency tables

#### 2.5 General form

---

### PROC CATMOD

Covariance Matrix of the Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Row</th>
<th>Parameter</th>
<th>Col1</th>
<th>Col2</th>
<th>Col3</th>
<th>Col4</th>
<th>Col5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Intercept 1</td>
<td>0.71222273</td>
<td>0.15396965</td>
<td>-0.21076475</td>
<td>-0.02084241</td>
<td>-0.04770286</td>
</tr>
<tr>
<td>2</td>
<td>Intercept 2</td>
<td>0.15396965</td>
<td>0.36882381</td>
<td>-0.02733780</td>
<td>-0.07988560</td>
<td>0.00053087</td>
</tr>
<tr>
<td>3</td>
<td>MALE</td>
<td>-0.21076475</td>
<td>-0.02733780</td>
<td>0.30752921</td>
<td>0.05217236</td>
<td>-0.04419596</td>
</tr>
<tr>
<td>4</td>
<td>MALE 2</td>
<td>-0.02084241</td>
<td>-0.07988560</td>
<td>0.05217236</td>
<td>0.13800928</td>
<td>0.01285102</td>
</tr>
<tr>
<td>5</td>
<td>BUSINESS 1</td>
<td>-0.04770286</td>
<td>0.00053087</td>
<td>-0.04419596</td>
<td>0.30100261</td>
<td>0.30100261</td>
</tr>
<tr>
<td>6</td>
<td>BUSINESS 2</td>
<td>-0.00080484</td>
<td>-0.0104324</td>
<td>-0.0285184</td>
<td>-0.02927663</td>
<td>0.08234738</td>
</tr>
<tr>
<td>7</td>
<td>PUNISH 1</td>
<td>-0.23078883</td>
<td>-0.06043407</td>
<td>0.03048402</td>
<td>0.00355595</td>
<td>0.00835542</td>
</tr>
<tr>
<td>8</td>
<td>PUNISH 2</td>
<td>-0.06556553</td>
<td>-0.13777275</td>
<td>0.01163384</td>
<td>0.00899338</td>
<td>0.00350680</td>
</tr>
<tr>
<td>9</td>
<td>EXPLAIN 1</td>
<td>-0.16613286</td>
<td>-0.06469989</td>
<td>-0.03954012</td>
<td>-0.0775333</td>
<td>0.04260608</td>
</tr>
<tr>
<td>10</td>
<td>EXPLAIN 2</td>
<td>-0.06474416</td>
<td>-0.14621530</td>
<td>-0.01389756</td>
<td>-0.0170087</td>
<td>-0.00698706</td>
</tr>
</tbody>
</table>

---

Covariance Matrix of the Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Row</th>
<th>Col6</th>
<th>Col7</th>
<th>Col8</th>
<th>Col9</th>
<th>Col10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.00080484</td>
<td>-0.23078883</td>
<td>-0.06566553</td>
<td>-0.16613286</td>
<td>-0.06474416</td>
</tr>
<tr>
<td>2</td>
<td>-0.00104324</td>
<td>-0.06043407</td>
<td>-0.13777275</td>
<td>-0.06469989</td>
<td>-0.14621530</td>
</tr>
<tr>
<td>3</td>
<td>-0.00825184</td>
<td>0.03048402</td>
<td>0.01163384</td>
<td>-0.03954012</td>
<td>-0.01389756</td>
</tr>
<tr>
<td>4</td>
<td>-0.02927663</td>
<td>0.00355595</td>
<td>0.00899338</td>
<td>-0.00775333</td>
<td>-0.01170087</td>
</tr>
<tr>
<td>5</td>
<td>0.00835542</td>
<td>-0.0350680</td>
<td>-0.04260608</td>
<td>-0.00698706</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.17921391</td>
<td>-0.00493934</td>
<td>-0.01127616</td>
<td>-0.0889248</td>
<td>-0.0128754</td>
</tr>
<tr>
<td>7</td>
<td>-0.00493934</td>
<td>0.11124149</td>
<td>0.03751659</td>
<td>0.03483635</td>
<td>0.01419912</td>
</tr>
<tr>
<td>8</td>
<td>-0.01127616</td>
<td>0.03751659</td>
<td>0.08346961</td>
<td>0.01072200</td>
<td>0.03133562</td>
</tr>
<tr>
<td>9</td>
<td>-0.00889248</td>
<td>0.03483635</td>
<td>0.01072200</td>
<td>0.29754745</td>
<td>0.07343716</td>
</tr>
<tr>
<td>10</td>
<td>-0.01287854</td>
<td>0.01419912</td>
<td>0.03133562</td>
<td>0.07343716</td>
<td>0.16273781</td>
</tr>
</tbody>
</table>
2. Unordered multinomial Logit model

2.1 Categorical and Multinomial distributions
2.2 Multinomial Logit model: basic idea
2.3 PROC CATMOD
2.4 Multinomial Logit model for contingency tables
2.5 General form

PROC CATMOD

- The DIRECT statement has the effect of retaining the quantitative variables as they are; if DIRECT is omitted, PUNISH, for example, will be converted into 2 dummy variables representing the 3 levels of PUNISH;

- The reference category is always the one with the highest value of the dependent variable;

- Note that Populations = 23 and Observations = 195.
# PROC CATMOD

- Coefficient estimates:

<table>
<thead>
<tr>
<th></th>
<th>Keep both vs Return both</th>
<th>Keep one vs Return both</th>
<th>Keep both vs Keep one</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-3.47**</td>
<td>-1.29**</td>
<td>-2.18</td>
</tr>
<tr>
<td>MALE</td>
<td>1.27*</td>
<td>1.17*</td>
<td>0.10</td>
</tr>
<tr>
<td>BUSINESS</td>
<td>1.18*</td>
<td>0.42</td>
<td>0.76</td>
</tr>
<tr>
<td>PUNISH</td>
<td>1.08**</td>
<td>0.20</td>
<td>0.89</td>
</tr>
<tr>
<td>EXPLAIN</td>
<td>-1.60**</td>
<td>-0.8*</td>
<td>-0.81</td>
</tr>
</tbody>
</table>

*: p-value < 0.05    **: p-value < 0.01

- Coefficient estimates in the second column have the same signs as but generally smaller than the corresponding estimates in the first column. Why?
How to work out the standard errors of $\hat{\beta}_{13}$, $\hat{\beta}_{23}$, $\hat{\beta}_{33}$ and $\hat{\beta}_{43}$?

Take $\hat{\beta}_{23}$ as an example. Note that $\hat{\beta}_{23} = \hat{\beta}_{21} - \hat{\beta}_{22}$. Hence

$$
\text{var}(\hat{\beta}_{23}) = \text{var}(\hat{\beta}_{21}) + \text{var}(\hat{\beta}_{22}) - 2\text{cov}(\hat{\beta}_{21}, \hat{\beta}_{22})
$$

$$
= 0.30752921 + 0.13800928 - 2(0.05217236)
$$

$$
= 0.341194
$$

and thus $s.e.(\hat{\beta}_{23}) = 0.584118$. 
Coefficient estimates and standard errors of estimates in the $\ln\left(\frac{p_{i1}}{p_{i2}}\right)$ equation:

<table>
<thead>
<tr>
<th></th>
<th>Keep both vs Keep one</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.18 (0.8773)*</td>
</tr>
<tr>
<td>MALE</td>
<td>0.10 (0.5841)</td>
</tr>
<tr>
<td>BUSINESS</td>
<td>0.76 (0.5619)</td>
</tr>
<tr>
<td>PUNISH</td>
<td>0.89 (0.3461)**</td>
</tr>
<tr>
<td>EXPLAIN</td>
<td>-0.81 (0.5593)</td>
</tr>
</tbody>
</table>

*: p-value $< 0.05$  **: p-value $< 0.01$
2.1 Categorical and Multinomial distributions
2.2 Multinomial Logit model: basic idea
2.3 PROC CATMOD
2.4 Multinomial Logit model for contingency tables
2.5 General form

PROC CATMOD

Interpretation of odds ratio estimates. Consider the 1st equation:

- For the MALE, $e^{1.27} = 3.56$, implying that the odds that males will keep both the money and the wallet rather than returning both are about 3.56 times the odds for females;

- For BUSINESS, $e^{1.27} = 3.56$, meaning the odds that a business student will keep both the money and the wallet rather than returning both are about 1.1804 times the odds for non-business students;

- For PUNISH, $e^{1.08} = 2.94$, meaning that each 1 level increase in PUNISH multiples the odds of keeping both vs. returning both by 2.94;

- For EXPLAIN, $e^{-1.60} = 0.25$, meaning that students whose parents explained their punishment had odds that were 25% the odds for those whose parents did not explain;
PROC CATMOD

- Wald test for the significance of an individual coefficient: as per binary Logit;
- Wald test for the joint significance of a subset of coefficients: for example, for the MALE,

\[ H_0 : \beta_{21} = \beta_{22} = 0 \quad \text{vs} \quad H_1 : \text{otherwise} \]

\[ W = 12.28 \sim \chi^2_2, \quad p - \text{value} = 0.0022 \]

Hence reject \( H_0 \).
PROC CATMOD

- Likelihood ratio (Deviance) test for overall significance of model:

\[ H_0 : \text{estimated and saturated models do not differ significantly, vs.} \]

\[ H_1 : \text{otherwise} \]

\[ LR = 2[lnL(\hat{\beta}_S) - lnL(\hat{\beta}_E)] \sim \chi^2_{2(p-k)}, \]

where \( p \) = number of population profile, and

\[ LR = 31.96, df = 2 \times (23 - 5) = 36, p-value = 0.6619 \]

Hence \( H_0 \) cannot be rejected, meaning that the estimated and saturated models do not differ significantly.
Consider the following contingency table:

<table>
<thead>
<tr>
<th>Preferred Automobile</th>
<th>Ford</th>
<th>Honda</th>
<th>Toyota</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locals</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>371</td>
<td>49</td>
<td>74</td>
</tr>
<tr>
<td>Male</td>
<td>250</td>
<td>45</td>
<td>71</td>
</tr>
<tr>
<td>Immigrants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>64</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>Male</td>
<td>25</td>
<td>5</td>
<td>13</td>
</tr>
</tbody>
</table>
Contingency table analysis

DATALINES;
1 1 1 371
1 1 2  49
1 1 3  74
1 0 1 250
1 0 2  45
1 0 3  71
0 1 1  64
0 1 2   9
0 1 3  15
0 0 1  25
0 0 2   5
0 0 3  13
;

PROC CATMOD DATA=automobile;
  WEIGHT freq;
  DIRECT local female;
  MODEL auto=local female / NOITER COVB;
RUN;
Contingency table analysis

The CATMOD Procedure

Data Summary

<table>
<thead>
<tr>
<th>Response</th>
<th>auto</th>
<th>Response Levels</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight Variable</td>
<td>freq</td>
<td>Populations</td>
<td>4</td>
</tr>
<tr>
<td>Data Set</td>
<td>AUTOMOBILE</td>
<td>Total Frequency</td>
<td>991</td>
</tr>
<tr>
<td>Frequency Missing</td>
<td>0</td>
<td>Observations</td>
<td>12</td>
</tr>
</tbody>
</table>

Population Profiles

<table>
<thead>
<tr>
<th>Sample</th>
<th>local</th>
<th>female</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>43</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>88</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>366</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>494</td>
</tr>
</tbody>
</table>

Response Profiles

<table>
<thead>
<tr>
<th>Response</th>
<th>auto</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Maximum Likelihood Analysis

Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2</td>
<td>34.14</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>local</td>
<td>2</td>
<td>2.08</td>
<td>0.3534</td>
</tr>
<tr>
<td>female</td>
<td>2</td>
<td>7.21</td>
<td>0.0272</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>2</td>
<td>0.85</td>
<td>0.6525</td>
</tr>
</tbody>
</table>
2. Unordered multinomial Logit model

2.1 Categorical and Multinomial distributions
2.2 Multinomial Logit model: basic idea
2.3 PROC CATMOD
2.4 Multinomial Logit model for contingency tables
2.5 General form

Contingency table analysis

---

The CATMOD Procedure

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Function</th>
<th>Parameter</th>
<th>Number</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intercept</td>
<td>1</td>
<td>0.8831</td>
<td>0.2426</td>
<td>13.25</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>-0.7580</td>
<td>0.3614</td>
<td>4.40</td>
<td>0.0359</td>
</tr>
<tr>
<td></td>
<td>local</td>
<td>1</td>
<td>0.3418</td>
<td>0.2370</td>
<td>2.08</td>
<td>0.1493</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.2710</td>
<td>0.3541</td>
<td>0.59</td>
<td>0.4442</td>
</tr>
<tr>
<td></td>
<td>female</td>
<td>1</td>
<td>0.4186</td>
<td>0.1713</td>
<td>5.97</td>
<td>0.0145</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.1051</td>
<td>0.2465</td>
<td>0.18</td>
<td>0.6700</td>
</tr>
</tbody>
</table>

Covariance Matrix of the Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Row</th>
<th>Parameter</th>
<th>Col1</th>
<th>Col2</th>
<th>Col3</th>
<th>Col4</th>
<th>Col5</th>
<th>Col6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Intercept</td>
<td>0.05887278</td>
<td>0.04441768</td>
<td>-0.04932291</td>
<td>-0.03735759</td>
<td>-0.01845122</td>
<td>-0.01423423</td>
</tr>
<tr>
<td>2</td>
<td>Intercept</td>
<td>0.04441768</td>
<td>0.13057697</td>
<td>-0.03739954</td>
<td>-0.11121948</td>
<td>-0.01416023</td>
<td>-0.03771544</td>
</tr>
<tr>
<td>3</td>
<td>local</td>
<td>-0.04932291</td>
<td>-0.03739954</td>
<td>0.05618380</td>
<td>0.04299291</td>
<td>0.00328073</td>
<td>0.00256331</td>
</tr>
<tr>
<td>4</td>
<td>local</td>
<td>-0.03735759</td>
<td>-0.11121948</td>
<td>0.04299291</td>
<td>0.12540426</td>
<td>0.00256331</td>
<td>0.00653709</td>
</tr>
<tr>
<td>5</td>
<td>female</td>
<td>-0.01845122</td>
<td>-0.01416023</td>
<td>0.00328073</td>
<td>0.00256331</td>
<td>0.02932694</td>
<td>0.02334582</td>
</tr>
<tr>
<td>6</td>
<td>female</td>
<td>-0.01423423</td>
<td>-0.03771544</td>
<td>0.00264707</td>
<td>0.00653709</td>
<td>0.02334582</td>
<td>0.06076618</td>
</tr>
</tbody>
</table>
Class exercise

1. Express the three estimated models in terms of log odds.
2. Compute the standard errors of the estimates in the third equation and determine their significance.
3. Test for the joint significance of the LOCAL coefficients.
4. Test for the joint significance of the FEMALE coefficients.
5. Test for the overall significance of the model.
6. In view of the statistical results, suggest a method to simplify the model.
General form

Let $p_{ij}$ be the probability that observation $i$ falls in category $j$, $j = 1, 2, \cdots, J$. The model is then

$$\ln\left(\frac{p_{ij}}{p_{ij}}\right) = Z_{ij}; \quad j = 1, 2, \cdots, J - 1$$

where $Z_{ij} = \beta_{1j} + \beta_{2j}X_{2i} + \beta_{3i}X_{3i} + \cdots \beta_{ki}X_{ki}$.

Note that each of the $J - 1$ categories is compared with the highest category $J$. 
These equations can be solved to yield

\[ p_{ij} = \frac{e^{Z_{ij}}}{1 + \sum_{t=1}^{J-1} e^{Z_{it}}}, \quad t = 1, 2, \ldots, J - 1, \]

and

\[ p_{iJ} = \frac{1}{1 + \sum_{t=1}^{J-1} e^{Z_{it}}}. \]
Ordered Logit model

- For data of ordinal nature, the unordered multinomial Logit model ignores an important piece of information in the data, namely, the ordering of the values;
- In the Wallet example, there is an obvious ordering in the responses, with 1 being the most unethical response, 3 the most ethical, and 2 the intermediate.
- Usually, the Ordered multinomial Logit model is simply referred to as the Ordered Logit model or Cumulative Logit model.
Latent variable

To see how we can deal with such data sensibly, let us reconsider the binary Logit set-up. One way to proceed is to assume that there exists an unobservable "latent" variable, $y^*$. The observed variable is coded $y$ such that

$$
y = \begin{cases} 
1 & \text{if } y^* > \mu \\
0 & \text{if } y^* \leq \mu
\end{cases}
$$

where $\mu$ is a threshold value.

For example, $y^*$ may be represented by a hidden utility function. If the satisfaction level exceeds $\mu$, then the restaurant is considered to be acceptable and $y$ takes on 1; otherwise it is considered to be unacceptable and $y$ takes on 0.
Latent variable

Let $y^* = \beta X + \epsilon$, where $X$ is a factor affecting $y^*$. In general, $X$ can be multivariate.

Then we have

$$
Pr(y = 1|X) = Pr(y^* > \mu|X) \\
= Pr(\beta X + \epsilon > \mu|X) \\
= Pr(\epsilon > \mu - \beta X|X) \\
= 1 - F(\mu - \beta X|X) \\
= F(\beta X - \mu|X),
$$

assuming a symmetric underlying distribution.
Setting $\mu = 0$, we have

$$Pr(y = 1|X) = F(\beta X|X)$$
$$Pr(y = 0|X) = 1 - F(\beta X|X)$$

and then choose the c.d.f. of the Logistic distribution as the link function.

Clearly, this results in the binary Logit model.
Now let us extend this framework to ordered data with $J$ categories. Let

$$y = \begin{cases} 
1 & \text{if } y^* \leq \mu_1 \\
2 & \text{if } \mu_1 < y^* \leq \mu_2 \\
3 & \text{if } \mu_2 < y^* \leq \mu_3 \\
\vdots \\
J & \text{if } \mu_{J-1} \leq y^*,
\end{cases}$$

where $\mu_1, \ldots, \mu_{J-1}$ are unknown thresholds to be estimated along with the $\beta$’s.
Thus,

\[
Pr(y = 1 | X) = Pr(\beta X + \epsilon \leq \mu_1 | X)
\]
\[
= Pr(\epsilon \leq \mu_1 - \beta X | X)
\]
\[
= F(\mu_1 - \beta X | X),
\]

\[
Pr(y = 2 | X) = Pr(\mu_1 < \beta X + \epsilon \leq \mu_2 | X)
\]
\[
= Pr(\mu_1 - \beta X < \epsilon \leq \mu_2 - \beta X | X)
\]
\[
= F(\mu_2 - \beta X | X) - F(\mu_1 - \beta X | X)
\]

and so on for \( y = 3, \ldots, J - 1 \).
Latent variable

\[ Pr(y = J | X) = Pr(\mu_{J-1} \leq \beta X + \epsilon | X) \]
\[ = Pr(\epsilon \geq \mu_{J-1} - \beta X | X) \]
\[ = 1 - Pr(\epsilon < \mu_{J-1} - \beta X | X) \]
\[ = 1 - F(\mu_{J-1} - \beta X | X), \]

where it is required that \( \mu_2 < \mu_2 < \mu_3 < \ldots \mu_{J-1} \) in order for these probabilities to be positive.
As for the effects exerted by the signs of the coefficients, we need to look carefully at the marginal effects. Recall that

\begin{align*}
p_1 &= F(\mu_1 - \beta X) \\
p_2 &= F(\mu_2 - \beta X) - F(\mu_1 - \beta X) \\
&\vdots \\
p_J &= 1 - F(\mu_{J-1} - \beta X)
\end{align*}
3.1 Basic idea and latent variable

Latent variable

- So for the marginal effects,

\[
\frac{\partial p_1}{\partial X} = -\beta f(\mu_1 - \beta X)
\]

\[
\frac{\partial p_j}{\partial X} = \beta(f(\mu_j - \beta X) - f(\mu_{j-1} - \beta X)), \ j = 2, \ldots, J - 1,
\]

\[
\frac{\partial p_J}{\partial X} = \beta f(\mu_{J-1} - \beta X)
\]
Latent variable

- So, the sign of the $p_1$ marginal effect is opposite to the coefficient sign, the sign of the $p_J$ marginal effect is the same as the coefficient sign, but other signs are ambiguous;
So, the sign of the $p_1$ marginal effect is opposite to the coefficient sign, the sign of the $p_J$ marginal effect is the same as the coefficient sign, but other signs are ambiguous;

One has to be very careful when interpreting the signs of the coefficients in this model.
Consider the Wallet example of three categories, $WALLET = 1, 2, 3$, and use the Logit as the link function. That is,

\[ p_1 = \frac{1}{1 + e^{\mu_1 - \beta X}} \]
\[ p_2 = \frac{1}{1 + e^{\mu_2 - \beta X}} - \frac{1}{1 + e^{\mu_1 - \beta X}} \]
\[ p_3 = 1 - \frac{1}{1 + e^{\mu_2 - \beta X}} \]
Consider the Wallet example of three categories, \( WALLET = 1, 2, 3 \), and use the Logit as the link function. That is,

\[
\begin{align*}
    p_1 &= \frac{1}{1 + e^{\mu_1 - \beta X}} \\
    p_2 &= \frac{1}{1 + e^{\mu_2 - \beta X}} - \frac{1}{1 + e^{\mu_1 - \beta X}} \\
    p_3 &= 1 - \frac{1}{1 + e^{\mu_2 - \beta X}}
\end{align*}
\]

The corresponding log of the odds equations are:

\[
\begin{align*}
    \ln\left( \frac{p_1}{p_2 + p_3} \right) &= Z_1 = \mu_1 - \beta X, \\
    \ln\left( \frac{p_1 + p_2}{p_3} \right) &= Z_2 = \mu_2 - \beta X
\end{align*}
\]
Estimation

DATA WALLET;
INFILE 'D:\TEACHING\MS4225\WALLET.TXT';
INPUT WALLET MALE BUSINESS PUNISH EXPLAIN;
PROC LOGISTIC DATA=WALLET;
OUTPUT OUT=OUT1 PREDICTED=P;
MODEL WALLET=MALE BUSINESS PUNISH EXPLAIN;
RUN;
PROC PRINT DATA=OUT1;
RUN;
The equations being estimated are:

\[ \ln \left( \frac{p_{i1}}{p_{i2} + p_{i3}} \right) = Z_{i1} = \mu_1 - \beta_{21} \text{MALE}_i - \beta_{31} \text{BUSINESS}_i - \beta_{41} \text{PUNISH}_i - \beta_{51} \text{EXPLAIN}_i; \]

\[ \ln \left( \frac{p_{i1} + p_{i2}}{p_{i3}} \right) = Z_{i2} = \mu_2 - \beta_{22} \text{MALE}_i - \beta_{32} \text{BUSINESS}_i - \beta_{42} \text{PUNISH}_i - \beta_{52} \text{EXPLAIN}_i; \]

The ordered Logit model imposes the following constraints:

\[ \beta_{21} = \beta_{22}, \beta_{31} = \beta_{32}, \beta_{41} = \beta_{42} \quad \text{and} \quad \beta_{51} = \beta_{52} \]
3.1 Basic idea and latent variable
3.2 Estimation

The LOGISTIC Procedure

Model Information

Data Set WORK.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

Response Profile

<table>
<thead>
<tr>
<th>Ordered Value</th>
<th>WALLET</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>121</td>
</tr>
</tbody>
</table>

Probabilities modeled are cumulated over the lower Ordered Values.

Model Convergence Status

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

Score Test for the Proportional Odds Assumption

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1514</td>
<td>4</td>
<td>0.2721</td>
</tr>
</tbody>
</table>

Model Fit Statistics

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Intercept Only</th>
<th>Intercept and Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>356.140</td>
<td>319.367</td>
</tr>
<tr>
<td>SC</td>
<td>362.686</td>
<td>339.005</td>
</tr>
<tr>
<td>-2 Log L</td>
<td>352.140</td>
<td>307.367</td>
</tr>
</tbody>
</table>
## 3. Ordered multinomial Logit model

### 3.1 Basic idea and latent variable

### 3.2 Estimation

**Testing Global Null Hypothesis: BETA=0**

<table>
<thead>
<tr>
<th>Test</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>44.7727</td>
<td>4</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Score</td>
<td>40.8753</td>
<td>4</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Wald</td>
<td>38.5746</td>
<td>4</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

**Analysis of Maximum Likelihood Estimates**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept 1</td>
<td>1</td>
<td>-3.2691</td>
<td>0.5612</td>
<td>33.9325</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Intercept 2</td>
<td>1</td>
<td>-1.4913</td>
<td>0.5085</td>
<td>8.6012</td>
<td>0.0034</td>
</tr>
<tr>
<td>MALE</td>
<td>1</td>
<td>1.0636</td>
<td>0.3255</td>
<td>10.6771</td>
<td>0.0011</td>
</tr>
<tr>
<td>BUSINESS</td>
<td>1</td>
<td>0.7370</td>
<td>0.3515</td>
<td>4.3973</td>
<td>0.0360</td>
</tr>
<tr>
<td>PUNISH</td>
<td>1</td>
<td>0.6874</td>
<td>0.2246</td>
<td>9.3644</td>
<td>0.0022</td>
</tr>
<tr>
<td>EXPLAIN</td>
<td>1</td>
<td>-1.0452</td>
<td>0.3392</td>
<td>9.4972</td>
<td>0.0021</td>
</tr>
</tbody>
</table>

**Odds Ratio Estimates**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALE</td>
<td>2.897</td>
<td>1.531 5.483</td>
</tr>
<tr>
<td>BUSINESS</td>
<td>2.090</td>
<td>1.049 4.161</td>
</tr>
<tr>
<td>PUNISH</td>
<td>1.989</td>
<td>1.280 3.089</td>
</tr>
<tr>
<td>EXPLAIN</td>
<td>0.352</td>
<td>0.181 0.684</td>
</tr>
</tbody>
</table>

**Association of Predicted Probabilities and Observed Responses**

<table>
<thead>
<tr>
<th>Percent Concordant</th>
<th>Somers' D</th>
<th>0.465</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent Discordant</td>
<td>Gamma</td>
<td>0.509</td>
</tr>
<tr>
<td>Percent Tied</td>
<td>Tau-a</td>
<td>0.250</td>
</tr>
<tr>
<td>Pairs</td>
<td>c</td>
<td>0.733</td>
</tr>
</tbody>
</table>
3. Ordered multinomial Logit model

3.1 Basic idea and latent variable

3.2 Estimation

### Estimation

<table>
<thead>
<tr>
<th>Obs</th>
<th>WALLET</th>
<th>MALE</th>
<th>BUSINESS</th>
<th>PUNISH</th>
<th>EXPLAIN</th>
<th><em>LEVEL</em></th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0.13076</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0.47091</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0.05024</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0.23836</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.02591</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0.13597</td>
</tr>
</tbody>
</table>
PROcedure: PROC LOGISTIC produces the following results:

\[
\ln\left(\frac{\hat{p}_{i1}}{p_{i2} + p_{i3}}\right) = \hat{Z}_{i1} = -3.2691 - (-1.0636)MALE_i - (-0.7370)BUSINESS_i - (-0.6874)PUNISH_i - (+1.0452)EXPLAIN_i
\]

\[
\ln\left(\frac{\hat{p}_{i1} + \hat{p}_{i2}}{p_{i3}}\right) = \hat{Z}_{i2} = -1.4913 - (-1.0636)MALE_i - (-0.7370)BUSINESS_i - (-0.6874)PUNISH_i - (+1.0452)EXPLAIN_i
\]
Hence we have, for example, the following marginal effects with respect to BUSINESS

\[
\frac{\partial \hat{p}_{i1}}{\partial \text{BUSINESS}_i} = -(-0.7370)f(\hat{Z}_{i1}),
\]

\[
\frac{\partial \hat{p}_{i2}}{\partial \text{BUSINESS}_i} = -0.7370(f(\hat{Z}_{i2}) - f(\hat{Z}_{i1})),
\]

\[
\frac{\partial \hat{p}_{i3}}{\partial \text{BUSINESS}_i} = -0.0730f(\hat{Z}_{i2})
\]

That is, other things being equal, a change in major from a non-business to a business related discipline will encourage student \(i\) to keep both the wallet and money and discourage him to return both.
However, it is uncertain how this change in major will impact his probability of keeping only the money and returning the wallet, as the change in probability of this latter event depends on $f(\hat{Z}_{i2}) - f(\hat{Z}_{i1})$, whose sign varies across $i$;

From output, for obs 1, $\hat{p}_{11} = 0.13076$ and $\hat{p}_{11} + \hat{p}_{12} = 0.47091$. Hence $\hat{p}_{12} = 0.34015$ and $\hat{p}_{13} = 0.529094$.

The interpretations of other results, e.g., LR, LM and Wald tests, odds ratio estimates, are analogous to those under binary Logit analysis.
The ordered Logit constrains the slope coefficients in the two log odds equations to be the same but allows the intercepts to differ. The ”Score test of the proportional odds assumption” tests if this is valid.

\[ H_0 : \beta_{21} = \beta_{22}, \beta_{31} = \beta_{32}, \beta_{41} = \beta_{42}, \beta_{51} = \beta_{52} \]

\[ H_1 : \text{otherwise} \]

\[ df = \text{number of restrictions under } H_0. \]
Estimation

▶ The ordered Logit constrains the slope coefficients in the two log odds equations to be the same but allows the intercepts to differ. The "Score test of the proportional odds assumption" tests if this is valid.

\[ H_0 : \beta_{21} = \beta_{22}, \beta_{31} = \beta_{32}, \beta_{41} = \beta_{42}, \beta_{51} = \beta_{52} \]

\[ H_1 : \text{otherwise} \]

\[ df = \text{number of restrictions under } H_0. \]

▶ Score = 5.1514, p-value = 0.2721. Hence we cannot reject \( H_0 \).