

DSC 410/510 Multivariate Statistical Methods

Discrete Choice Analysis

1

What is Discrete Choice Analysis?

- Used to evaluate consumer choice
 - ◆ Conjoint analysis: consumers evaluate product preferences by ranking combinations of attribute levels
 - ◆ Discrete choice analysis: each consumer chooses one combination of attribute levels from a set of attribute/level combinations
- Set of attribute/level combinations is a *choice set*
 - ◆ Consumers asked to choose one combination from each of several different choice sets
- Product attributes defining the choices are *choice attributes* as opposed to *non-choice attributes* that may be of interest but do not define the choices (e.g. demographics related to the consumers)

2

Discrete Choice Analysis Questions

- Which combination of attribute levels is most likely to be chosen by consumers?
- Which combination of attribute levels is least likely to be chosen by consumers?
- What is the probability of my product being chosen by a consumer?
- How does each attribute level influence the probability of choice?
- Do consumer characteristics influence choice?

3

Example: Chocolates

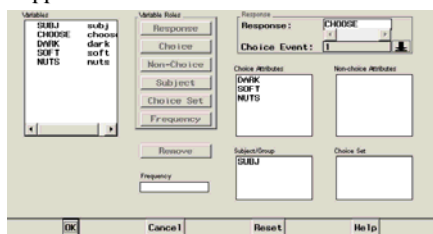
- Data are choices by ten subjects of one of eight chocolates
- Eight chocolates in the choice set are combinations of three choice attributes:
 - ◆ dark, soft, nuts
- Response variable is “choose” with 1 representing a choice and 0 representing a non-choice
 - ◆ so one 1 and seven 0’s for each subject

4

SAS: Chocolates

- Data: `chocs.xls`
- SAS: market research application

subj	choose	dark	soft	nuts
1	0	0	0	0
1	0	0	0	1
1	0	0	1	0
1	0	0	1	1
1	1	1	0	0
1	0	1	0	1
1	0	1	1	0
1	0	1	1	1
2	0	0	0	0



5

Multinomial Logit Estimation

- Suppose there are three alternatives: A, B, and C
- Each alternative is described by a vector X of k predictors (could include dummy variables for brands, continuous or discrete price variables, or other nominal or continuous attributes)
- Multinomial logit solves for a vector β of k parameters using the method of maximum likelihood: for any alternative within any set, Utility, $U = \exp(V) = \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)$

6

Multinomial Logit Probabilities

- Probability that alternative A is chosen given its attributes and those of the other two alternatives is
 - ◆ $pA| [A,B,C] = U_A / (U_A + U_B + U_C)$
- Similarly $pB| [A,B,C] = U_B / (U_A + U_B + U_C)$ and $pC| [A,B,C] = U_C / (U_A + U_B + U_C)$
 - ◆ Note that $pA + pB + pC = 1$
- Model parameters are best interpreted by exponentiating (called “odds or hazard ratios”):
 - ◆ e.g. if X_1 is a dummy variable indicating presence of an attribute, then $\exp(\beta_1) =$ choice probability when that attribute is present ÷ choice probability when that attribute is absent

7

SAS Output

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square
dark	1	1.38629	0.79057	3.0749
soft	1	-2.19722	1.05409	4.3450
nuts	1	0.84730	0.69007	1.5076

Pr > Chi-Square	Hazard Ratio	Label
0.0795	4.000	dark
0.0371	0.111	soft
0.2195	2.333	nuts

8

Interpretation

- $U = \exp(1.38629 \text{ dark} - 2.19722 \text{ soft} + 0.84730 \text{ nuts})$
 - ◆ e.g. $U_{(1,0,1)} = \exp(1.38629 + 0.84730) = 9.333$
 - ◆ see **chocsres.xls** spreadsheet
- Sum of the utilities = 18.518
- Choice probability = $U/18.518$
 - ◆ e.g. $9.333/18.518 = 0.504$
- Hazard ratio for “dark” is $\exp(1.38629) = 4.000$
 - ◆ i.e. presence of the “dark” attribute increases the odds of choice by 4 times

9

SAS: Response Variable

- Response variable contains values which indicate whether the choice alternative represented by an observation was chosen or not
- After selecting a response variable, you must select the choice value:
 - ◆ usually 1 indicates that the choice alternative represented by an observation was chosen, 0 indicates the choice alternative was not chosen
 - ◆ you must indicate which value is used to indicate a choice: select the down arrow to the right of the choice value field and select one value from the ensuing list

10

SAS: Choice Attributes

- Choice attributes are the variables that define the choice alternatives
 - ◆ in particular, variables used for the descriptions of the alternatives in “stated preference” experiments
 - ◆ or, differentiating attributes of actual alternatives in “revealed preference” experiments
- Each choice attribute variable must be numeric
 - ◆ so if some of the choice attributes are qualitative then dummy variables are needed

11

SAS: Non-choice Attributes

- Non-choice attributes are predictor (independent) variables that do not define the choice alternatives
 - ◆ e.g. characteristics of the respondents
- Choice attributes and non-choice attributes actually get analyzed in the same way ...
 - ◆ i.e. they are the X-variables from slide 6
- But you have to be a little careful with exactly how they enter the model (often in interactions)
 - ◆ e.g. see **travel.xls** dataset (slide 18)

12

SAS: Subject Variable

- The subject variable distinguishes different respondents or groups of respondents
 - ◆ e.g. in the chocolates dataset each row corresponds to a choice alternative for each respondent (so there are eight rows for each respondent, and these eight rows all have a common subject number)
- You need a subject variable if you do NOT designate a frequency variable – see next slide

13

SAS: Frequency Variable

- Discrete choice data can have a different form in which each row of the dataset corresponds to a choice alternative, but a frequency variable indicates how many respondents made that particular choice in the experiment
 - ◆ e.g. see `price.xls` dataset (slides 19-20)
- Each observation is treated as if it occurred n times where n is the value of the frequency variable for that observation
- If a frequency variable is selected, a subject variable is not necessary

14

SAS: Choice Set Variable

- The choice set variable contains values that distinguish different sets of choice alternatives
- A respondent makes one choice from among the alternatives in each choice set
 - ◆ e.g. there was only one choice set for the chocolates example, so there was no need for a choice set variable
 - ◆ e.g. there are eight choice sets for the price example, so the choice set variable takes on the values 1-8

15

SAS: Results

- Bar chart displays significances of each of the choice and non-choice attributes
- Statistics assessing the overall fit of the model can be displayed, including the log-likelihood, the Score test, and the Wald test
- A chi-square statistic is displayed for all three tests with the associated degrees of freedom and p-value
- The model-fit statistics are given with and without covariates (the choice and non-choice attributes)
- Choice probabilities can be displayed

16

SAS: Parameter Estimates

- The significance of the individual parameter estimates allow you to determine which choice and non-choice attributes have an effect on making a choice
- The parameter estimate, degrees of freedom, standard error, Wald chi-square, p-value, and hazard ratio (exponentiated parameter estimate) are displayed for each choice and non-choice attribute
- See slides 8-9 for example of interpretation

17

Example: Travel

- 21 respondents asked to choose between travel by auto, plane, or public transit (bus or train)
- Choice possibilities defined by three attributes: travel time (continuous), travel mode (qualitative with 3 levels), and respondent age (continuous)
- Use Choose as the response with a choice value of 1 and Subj as the subject variable
- Auto, Plane, and Time are choice attributes
 - ◆ why not Transit?
- AgeAuto and AgePlane are non-choice attributes
 - ◆ why not Age or AgeTransit?

18

Example: Price/Brand

- 100 respondents presented with brands of a product at different prices
- Choice possibilities defined by two attributes: price and brand
- Five choice alternatives presented at a time to a respondent, from which one alternative is chosen
- Eight of these choice sets were presented, each one with a different set of five combinations of price and brand

19

Example: Price/Brand

- The brand/price combinations for the first two choice sets are:

Set	Brand1	Brand2	Brand3	Brand4	Other
1	\$3.99	\$3.99	\$3.99	\$3.99	\$4.99
2	\$3.99	\$3.99	\$5.99	\$5.99	\$4.99
- Use the variable Choose as the response with a choice value of 1 and Freq as the frequency variable
- Use Price and dummy-coded Brand1-Brand4 as choice attributes and Set as a choice set variable

20

Multinomial Logit Drawbacks

- Independence from irrelevant alternatives (IIA): odds of choosing one alternative over another do not depend on other alternatives in the choice set
 - ◆ Unrealistic sometimes, e.g. for travel options Auto and Red Bus suppose 80% choose Auto (i.e. odds of 4 to 1)
 - ◆ Now suppose options are Auto, Red Bus and Blue Bus – expect odds for choosing Auto over Red Bus to be 8 to 1 (since 80% will still choose Auto, and presumably about 10% each will choose Red Bus and Blue Bus), but IIA says the odds remain 4 to 1
- Multinomial logit does not explicitly account for correlations in unobserved utility over repeated choices by each respondent

21

Multinomial Logit Extensions

- Multinomial logit models with random effects
 - ◆ Also referred to as mixed, random-parameters, or error-components logit models
 - ◆ Each respondent has their own “random effect” which comes from some common distribution
 - ◆ Possible to apply using SAS (Proc NLMIXED) or from a Bayesian perspective with specialized software (e.g. Sawtooth Software, WinBUGS)
- Ordinal responses (naturally ordered categories)
 - ◆ e.g. see “Measuring Consumer Preference for Socially Responsible Products” paper at <http://1cb1.uoregon.edu/ipardoe/research.htm>

22