Course: Customer Analytics

CHOICE ANALYSIS

ANA ALINA TUDORAN

DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

AGENDA

I. Discuss the objective and design of choice analysis
II. Run the application in R and interpret
III. Hand-on application with new data

OBJECTIVE

- To assess consumer preferences
- To understand how features (e.g., package size, brand, flavor, price, privacy, ...) affect which product a customer will choose
- To understand how they trade off desirable features against each other
- The outcome (DV) is not a number, but a product choice among several options

SCENARIO FRAMING

- **Setting:** A company designs a new line of minivans with different characteristics
- **Research problem:** How large the minivan should be and what type of engine it should have?
- **Objective:** The experiment/survey aims to identify the factors most customers desire. For this experiment, the company selects five main characteristics (factors or numeric):
    - Carpool
    - Seat
    - Cargo
    - Engine
    - Price

FULL FACTORIAL DESIGN

Factors:	**Levels:**	**Max product profiles**
1. Carpool	yes, no	2*3*2*3*3 = 108
2. Seat	6,7,8	
3. Cargo	2ft; 3ft	
4. Engine	gas, hyb, elec	
5. Price	30, 35, 40	

SURVEY QUESTION

Each respondent answers several questions, where options have varying attributes.

Which of the following minivans would you buy?
Assume all three minivans are identical other than the features listed below.

	Option 1	Option 2	Option 3
	6 passengers 2 ft. cargo area gas engine $35,000	8 passengers 3 ft. cargo area hybrid engine $30,000	6 passengers 3 ft. cargo area gas engine $30,000
I prefer (check one):	☐	☐	☑

Fig. 13.1 An example choice-based conjoint survey question

SURVEY QUESTION

Each respondent answers several questions, where options have varying attributes.

Which of the following minivans would you buy?
Assume all three minivans are identical other than the features listed below.

	Option 1	Option 2	Option 3
	6 passengers	8 passengers	6 passengers
	2 ft. cargo area	3 ft. cargo area	3 ft. cargo area
	gas engine	hybrid engine	gas engine
	$35,000	$30,000	$30,000
I prefer (check one):	☐	☐	☑

ALTERNATIVES

ATRIBUTES

LEVEL OF ATTRIBUTE

Fig. 13.1 An example choice-based conjoint survey question

SEVERAL QUESTIONS PER CUSTOMER

- Each question has the same structure, but varies the levels of the attributes for the alternatives
- In this study, each respondent sees 15 questions. (15 x3 = 45 product profiles out of 108 possible)
- A typical study might have 5–10 attributes and include 10–20 questions for each respondent
- The number of product profiles to choose from can also vary

FULL VS. PARTIAL FACTORIAL DESIGN

- FULL = Tests every possible combination of factors and levels
  - Comprehensive: Captures all possible interactions between factors.
  - Clarity: Provides clear and detailed information about the effects of each factor and their interactions.
  - Resource-Intensive: Requires a large number of experiments as the number of factors increases.
  - Time-consuming: More experimental runs mean more time needed for completion.

- PARTIAL /FRACTIONAL = Tests only a subset of the combinations of factors and levels
  - Efficient: Requires fewer experiments, saving time and resources
  - Practical: Ideal for preliminary studies or when resources are limited
  - May miss some interactions between factors.
  - The choice of which combinations to test must be strategic to ensure relevant data is collected.

METHODS OF ANALYSIS

- Choice-based conjoint analysis
  - Individuals choose among products => DV is categorical
  - Multinomial logit model (focus in this lecture)

- Rating-based (metric) conjoint analysis
  - Individuals give ratings to single products => DV is metric
  - Liner model

DATA STRUCTURE

- Depending on the software
- In R, long format

	resp.id	ques	alt	carpool	seat	cargo	eng	price	choice
1	1	1	1	yes	6	2ft	gas	35	0
2	1	1	2	yes	8	3ft	hyb	30	0
3	1	1	3	yes	6	3ft	gas	30	1
4	1	2	1	yes	6	2ft	gas	30	0
5	1	2	2	yes	7	3ft	gas	35	1
6	1	2	3	yes	6	2ft	elec	35	0
7	1	3	1	yes	8	3ft	gas	35	1
8	1	3	2	yes	7	3ft	elec	30	0
9	1	3	3	yes	8	2ft	elec	40	0
10	1	4	1	yes	7	3ft	elec	40	1
11	1	4	2	yes	8	3ft	gas	40	0
12	1	4	3	yes	6	2ft	hyb	30	0
13	1	5	1	yes	6	3ft	elec	30	1
14	1	5	2	yes	6	3ft	elec	40	0
15	1	5	3	yes	6	3ft	hyb	40	0
16	1	6	1	yes	6	2ft	elec	40	0
17	1	6	2	yes	7	2ft	gas	35	1
18	1	6	3	yes	7	2ft	hyb	35	0
19	1	7	1	yes	6	3ft	hyb	35	1
20	1	7	2	yes	7	3ft	gas	40	0
21	1	7	3	yes	6	2ft	elec	30	0
22	1	8	1	yes	6	3ft	elec	35	0
23	1	8	2	yes	8	3ft	hyb	40	0
24	1	8	3	yes	7	3ft	hyb	30	1

Showing 1 to 24 of 9,000 entries, 9 total columns

STEPS OF ANALYSIS IN R

1. Ensure correct data structure
2. Perform some descriptive statistics, e.g.
   - *Summary()* - How many times each level appear in the questions
   - *Xtabs()* - the number of times respondents chose an alternative at each feature level.
3. Perform multinomial logit model with “mlogit” package
   - It requires transforming data in a special format with *logit.data()* function
   - The syntax for setting the model is standard

RESULTS

- The estimated parameter (part worth coefficient) for each level, relative to the base levels of each attribute
- Customers disliked electric and hybrid engines (relative to the base level, which is gas) and disliked the $40K and $30K price (relative to the base level price of $30)
- All parameter estimates are on the logit scale and typically range between –2 and 2

```
Call:
mlogit(formula = choice ~ 0 + seat + cargo + eng + price, data = cbc.mlogit,
    method = "nr")

Frequencies of alternatives:choice
      1       2       3
0.32700 0.33467 0.33833

nr method
5 iterations, 0h:0m:0s
g'(-H)^-1g = 7.84E-05
successive function values within tolerance limits

Coefficients :
          Estimate Std. Error  z-value  Pr(>|z|)
seat7    -0.535280   0.062360  -8.5837 < 2.2e-16 ***
seat8    -0.305840   0.061129  -5.0032 5.638e-07 ***
cargo3ft  0.477449   0.050888   9.3824 < 2.2e-16 ***
enghyb   -0.811282   0.060130 -13.4921 < 2.2e-16 ***
engelec  -1.530762   0.067456 -22.6926 < 2.2e-16 ***
price35  -0.913656   0.060601 -15.0765 < 2.2e-16 ***
price40  -1.725851   0.069631 -24.7856 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log-Likelihood: -2581.6
```

MAIN REFERENCES

Chapman C. & Feit E. M. (2015). *R for Marketing Research and Analytics*, Springer. Chapter 13. Retrieved from https://soeg.kb.dk/permalink/45KBDK_KGL/1pioq0f/alma9912306264090576 3

Research Project:
https://docs.google.com/forms/d/e/1FAIpQLSdBgLnkUTZIiFAEx0FYHlcLT4zcz0vKf2gyvRJCMi2fL2R7qg/viewform?usp=sf_link