

Assembling choice data

CHOICE MODELING FOR MARKETING IN R



Elea McDonnell Feit




Assistant Professor of Marketing, Drexel University

Choices observed in the "wild"

- Purchases in the grocery store
- Purchases from an online store
- Viewing choices on a video streaming service
- Voting for political candidates
- Choice of a marriage partner

This is sometimes called "revealed preference" data.

Survey choices

Scenario 1 of 3	Vehicle 1	Vehicle 2	Vehicle 3
Styling			
AWD/FWD	All Wheel Drive (AWD)	Front Wheel Drive (FWD)	All Wheel Drive (AWD)
Fuel Economy	20 mpg city	16 mpg city	26 mpg city
Engine	4 cylinder hybrid	6 cylinder	4 cylinder
Seating	8 passengers	8 passengers	5 passengers
Cargo Capacity	35 Cu. Ft. (about 7 large suitcases)	35 Cu. Ft. (about 7 large suitcases)	35 Cu. Ft. (about 7 large suitcases)
Max Cargo Capacity (seats folded down)	small (60 Cu. Ft.)	small (60 Cu. Ft.)	small (60 Cu. Ft.)
Price (MSRP)	\$24,999	\$24,999	\$27,999
Which of these vehicles would you be most likely to buy?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

This is called "conjoint data" or "stated preference data".

Long format choice data

Each observation is described by three rows: one for each option

ques	alt	choice	seat	trans	price
1	1	0	2	manual	35
1	2	0	5	auto	40
1	3	1	5	auto	30
2	1	1	5	manual	35
2	2	0	2	manual	30
2	3	0	4	auto	35

Wide format choice data

Attributes are repeated for each alternative

In wide choice data, each row is a choice.

ques	alt	choice	seat.1	seat.2	seat.3	trans.1	trans.2	trans.3
1	1	3	2	5	5	manual	auto	40
2	1	1	5	2	4	manual	manual	30

Wide format choice data in R

```
head(sportscar_wide)
```

```
  resp_id ques segment choice seat.1 seat.2 seat.3
1        1   1  basic      3      2      5      5
2        1   2  basic      1      5      2      4
3        1   3  basic      1      5      4      4
  trans.1 trans.2 trans.3 convert.1 convert.2
1 manual    auto    auto      yes      no
2 manual manual    auto      no      no
3  auto    auto manual    yes      yes
```

```
nrow(sportscar_wide)
```

```
2000
```

What types of chocolate do people choose?

CHOICE MODELING FOR MARKETING IN R

Converting from wide to long

CHOICE MODELING FOR MARKETING IN R



Elea McDonnell Feit

Assistant Professor of Marketing, Drexel
University

Why long format?

Each observation is described by three rows: one for each option

ques	alt	choice	seat	trans	price
1	1	0	2	manual	35
1	2	0	5	auto	40
1	3	1	5	auto	30
2	1	1	5	manual	35
2	2	0	2	manual	30
2	3	0	4	auto	35

Sportscar data in wide format

```
head(sportscar_wide)
```

	resp_id	ques	segment	choice	seat.1	seat.2	seat.3	trans.1	trans.2	trans.3
1	1	1	basic	3	2	5	5	manual	auto	auto
2	1	2	basic	2	5	2	4	manual	manual	auto
3	1	3	basic	1	5	4	4	auto	auto	manual
4	1	4	basic	3	2	4	4	manual	manual	auto
5	1	5	basic	2	5	5	2	manual	manual	auto
6	1	6	basic	3	2	4	2	auto	manual	auto

	convert.1	convert.2	convert.3	price.1	price.2	price.3
1	yes	no	no	35	40	30
2	no	no	no	35	30	35
3	yes	yes	no	35	30	40
4	yes	yes	yes	30	40	35
5	yes	no	yes	40	30	40
6	yes	yes	no	35	35	30

Transforming from wide to long

```
sportscar <- reshape(
  sportscar_wide,
  direction = "long",
  varying = list(seat = 5:7, trans = 8:10, convert = 11:13, price = 14:16),
  v.names = c("seat", "trans", "convert", "price"),
  timevar = "alt")
head(sportscar)
```

	resp_id	ques	segment	choice	alt	seat	trans	convert	price	id
1.1	1	1	basic	3	1	2	manual	yes	35	1
2.1	1	2	basic	2	1	5	manual	no	35	2
3.1	1	3	basic	1	1	5	auto	yes	35	3
4.1	1	4	basic	3	1	2	manual	yes	30	4
5.1	1	5	basic	2	1	5	manual	yes	40	5
6.1	1	6	basic	3	1	2	auto	yes	35	6

Sorting the long data

```
new_order <- order(sportscar$resp_id,  
                  sportscar$ques,  
                  sportscar$alt)  
sportscar <- sportscar[new_order,]  
head(sportscar)
```

	resp_id	ques	segment	choice	alt	seat	trans	convert	price	id
1.1	1	1	basic	3	1	2	manual	yes	35	1
1.2	1	1	basic	3	2	5	auto	no	40	1
1.3	1	1	basic	3	3	5	auto	no	30	1
2.1	1	2	basic	1	1	5	manual	no	35	2
2.2	1	2	basic	1	2	2	manual	no	30	2
2.3	1	2	basic	1	3	4	auto	no	35	2

Converting choice to a logical

```
sportscar$choice <- sportscar$choice == sportscar$alt  
head(sportscar)
```

	resp_id	ques	segment	choice	alt	seat	trans	convert	price	id
1.1	1	1	basic	FALSE	1	2	manual	yes	35	1
1.2	1	1	basic	FALSE	2	5	auto	no	40	1
1.3	1	1	basic	TRUE	3	5	auto	no	30	1
2.1	1	2	basic	TRUE	1	5	manual	no	35	2
2.2	1	2	basic	FALSE	2	2	manual	no	30	2
2.3	1	2	basic	FALSE	3	4	auto	no	35	2

Let's practice!

CHOICE MODELING FOR MARKETING IN R

Choice data in two files

CHOICE MODELING FOR MARKETING IN R



Elea McDonnell Feit

Assistant Professor of Marketing, Drexel University

Choice data in two files

Alternatives in one file

```
sportscar_alts[1:3, ]
```

```
  resp_id ques alt seat  trans convert price
1       1   1   1   2 manual    yes    35
2       1   1   2   5   auto    no    40
3       1   1   3   5   auto    no    30
```

Choices in another file

```
sportscar_choices[1, ]
```

```
  resp_id ques segment choice
1       1   1   basic      3
```


Merging the two files

```
sportscar <- merge(sportscar_choices, sportscar_alts,  
                  by = c("resp_id", "ques"))  
  
sportscar
```

	resp_id	ques	segment	choice	alt	seat	trans	convert	price
1	1	1	basic	3	1	2	manual	yes	35
2	1	1	basic	3	2	5	auto	no	40
3	1	1	basic	3	3	5	auto	no	30
4	1	10	basic	1	1	5	auto	yes	40
5	1	10	basic	1	2	4	auto	no	30
6	1	10	basic	1	3	2	manual	yes	40

Let's practice!

CHOICE MODELING FOR MARKETING IN R

Visualizing choice data

CHOICE MODELING FOR MARKETING IN R



Elea McDonnell Feit

Assistant Professor of Marketing, Drexel University

```
xtabs(~ trans, data = sportscar)
```

```
trans
  auto manual
3001  2999
```

```
xtabs(~ trans + choice, data = sportscar)
```

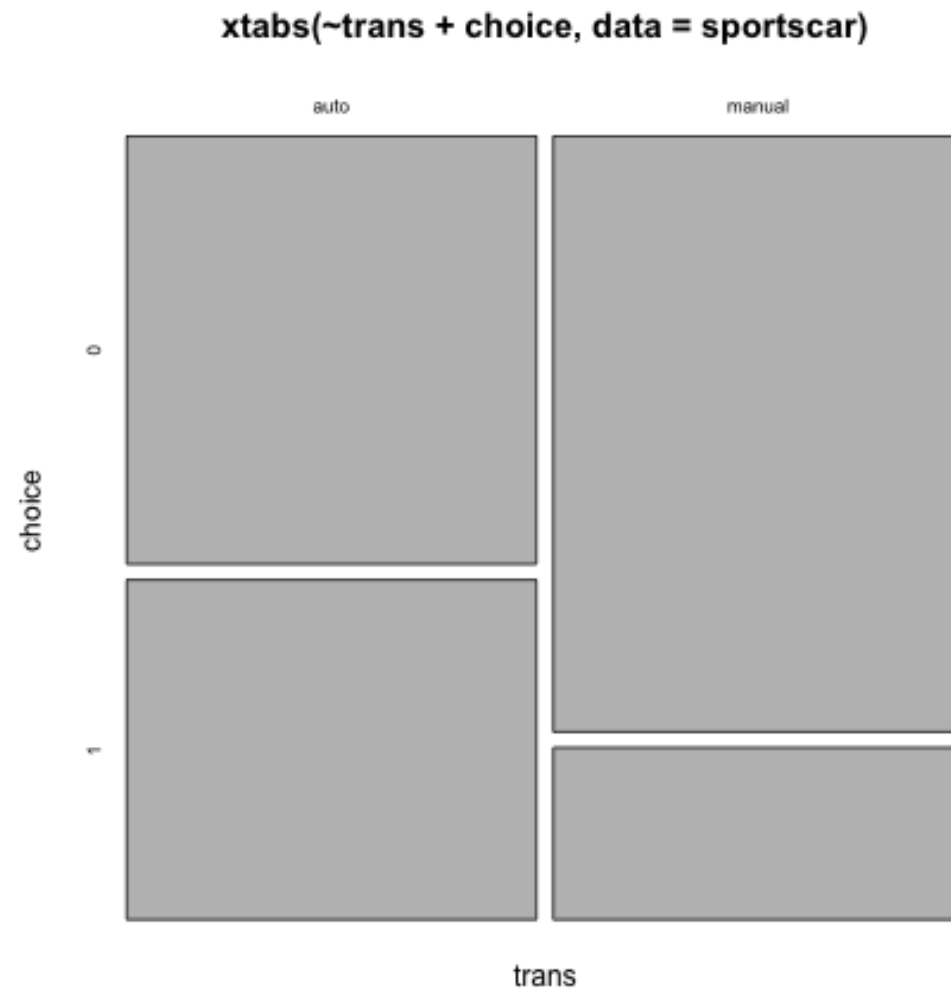
```
      choice
trans      0    1
  auto  1673 1328
 manual 2327  672
```

```
xtabs(choice ~ trans, data=sportscar)
```

```
trans
  auto manual
1328   672
```

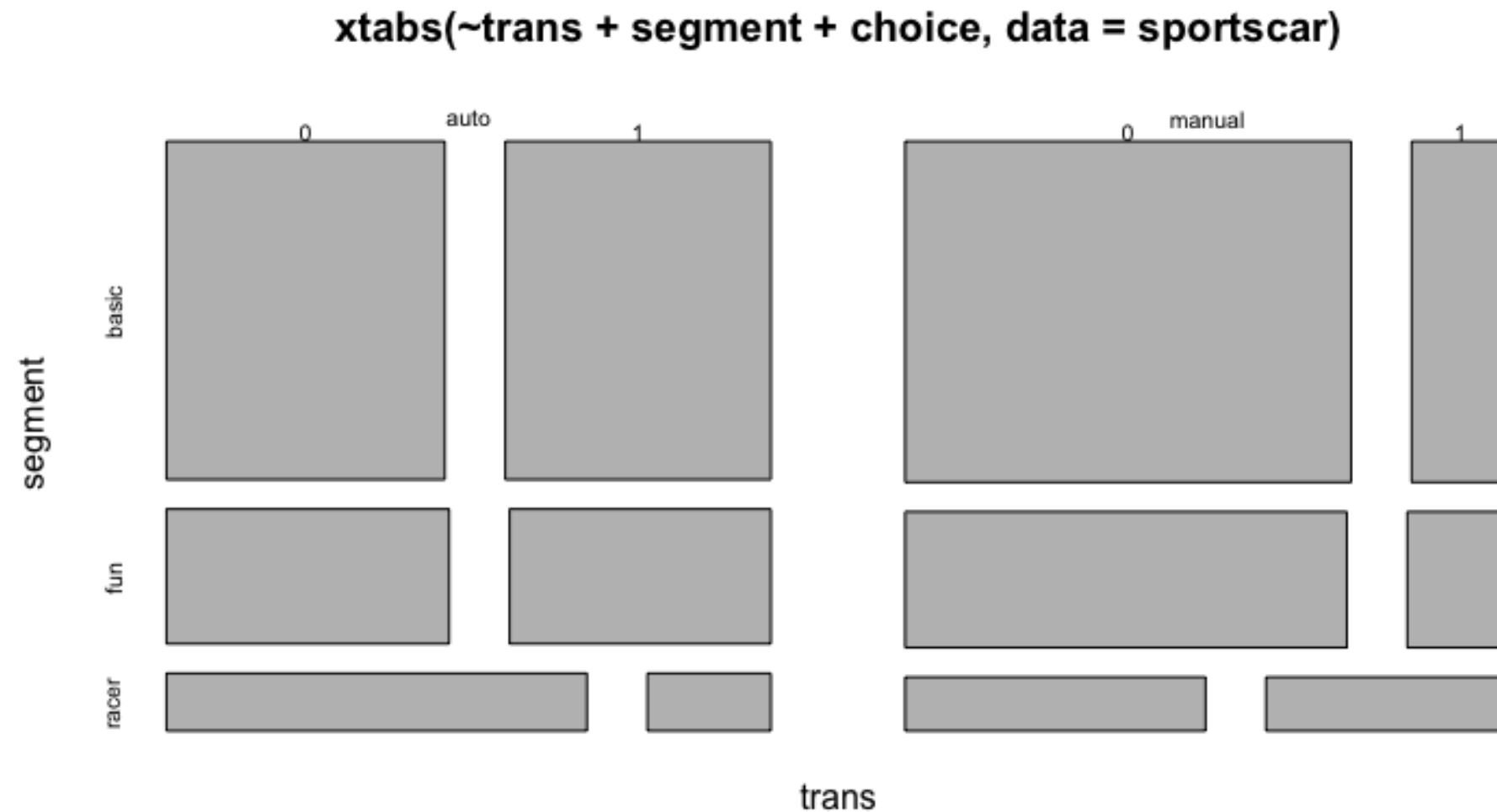
Plotting the output of xtabs()

```
plot(xtabs(~ trans + choice, data = sportscar))
```



Transmission choice by segment

```
plot(xtabs(~ trans + segment + choice, data = sportscar))
```



Let's practice!

CHOICE MODELING FOR MARKETING IN R

Designing a conjoint survey




CHOICE MODELING FOR MARKETING IN R



Elea McDonnell Feit

Assistant Professor of Marketing, Drexel University

Conjoint survey

Scenario 1 of 3	Vehicle 1	Vehicle 2	Vehicle 3
Styling			
AWD/FWD	All Wheel Drive (AWD)	Front Wheel Drive (FWD)	All Wheel Drive (AWD)
Fuel Economy	20 mpg city	16 mpg city	26 mpg city
Engine	4 cylinder hybrid	6 cylinder	4 cylinder
Seating	8 passengers	8 passengers	5 passengers
Cargo Capacity	35 Cu. Ft. (about 7 large suitcases)	35 Cu. Ft. (about 7 large suitcases)	35 Cu. Ft. (about 7 large suitcases)
Max Cargo Capacity (seats folded down)	small (60 Cu. Ft.)	small (60 Cu. Ft.)	small (60 Cu. Ft.)
Price (MSRP)	\$24,999	\$24,999	\$27,999
Which of these vehicles would you be most likely to buy?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Attributes and levels

Type

milk, dark, milk with nuts, dark with nuts, white

Brand

Dove, Ghirardelli, Godiva, Hershey's, Lindt

Price

0.5, 0.6, ... steps of 0.1 ..., 3.9, 4.0

Designing a choice survey

```
choc_survey[choc_survey$Subject == 1 & choc_survey$Trial == 1, ]
```

	Subject	Trial	Alt	Type	Brand	Price
1	1	1	1	NA	NA	NA
2	1	1	2	NA	NA	NA
3	1	1	3	NA	NA	NA

Creating a random design part 1

```
# Setup your attributes and levels list
attribs <- list(Type = c("Milk", "Dark", "White"),
                Brand = c("Cadbury", "Toblerone", "Kinder"),
                Price = 5:30 / 10)

# Create all possible combinations of attributes
all_comb <- expand.grid(attribs)
nrow(all_comb)
head(all_comb)
```

```
144
Type      Brand Price
1  Milk    Cadbury  0.5
2  Dark    Cadbury  0.5
3  White   Cadbury  0.5
4  Milk    Toblerone 0.5
5  Dark    Toblerone 0.5
6  White   Toblerone 0.5
```

Creating a random design part 2

```
for (i in 1:100) {  
  rand_rows <- sample(1:nrow(all_comb), size = 12 * 3)  
  rand_alts <- all_comb[rand_rows, ]  
  choc_survey[choc_survey$Subject == i, 4:6] <- rand_alts  
}
```

Fielding your survey: options

- Code up the survey yourself.
- Upload the survey design to a survey tool like Google Forms or Survey Monkey.
- Use a survey tool with a build-in conjoint design feature like Sawtooth, Conjoint.ly or Qualtrics.

Go field a conjoint survey!

CHOICE MODELING FOR MARKETING IN R