Experimental Design and Data Issues for Evaluating Soft Variables

Introduction to generating designs for qualitative variables

John Rose |Prof.
Jan 22 2012

THE UNIVERSITY OF SYDNEY

Handling soft (qualitative variables)

1

Qual versus quant attributes

- Unlike quantitative attributes, the implicit assumption is that the marginal utility of a qualitative attribute is non-linear with respect to the levels of the attribute
  - Why such an assumption is not implicit for quantitative attributes is an open question
- Consider the following experiment

Please consider a situation where you have the opportunity to go out on a date with two people who can be described as follows. Please look at the description of each person and select the person you would be more likely to date.

	Person A	Person B	None
Education	Post Grad	High School	
Employment	Unemployed	White collar	
Personality	Extrovert	Introvert	
Looks	Above average	Average	
I would choose			

Levels		
High School	Undergrad	Post grad
Unemployed	Blue collar	White collar
Introvert	Extrovert	
Below Average	Average	Above Average

Coding

› We could use linear coding for this such that

	Person A	Person B	None
Education	Post Grad	High School	
Employment	Unemployed	White collar	
Personality	Extrovert	Introvert	
Looks	Above average	Average	
I would choose			

	Levels	
High School	Undergrad	Post grad
Unemployed	Blue collar	White collar
Introvert	Extrovert	
Below Average	Average	Above Average

Code	Education	Employment	Personality	Looks
0	High School	Unemployed	Introvert	Below Average
1	Undergrad	Blue collar	Extrovert	Average
2	Post grad	White collar		Above Average

› In this case, the utility function would look

$$V_j = \alpha_j + \beta_{ed} ed + \beta_{Em} em + \beta_{pe} pe + \beta_{lo} lo$$

Dummy coding

› We could use linear coding for this such that

	Person A	Person B	None
Education	Post Grad	High School	
Employment	Unemployed	White collar	
Personality	Extrovert	Introvert	
Looks	Above average	Average	
I would choose			

Levels		
High School	Undergrad	Post grad
Unemployed	Blue collar	White collar
Introvert	Extrovert	
Below Average	Average	Above Average

Code	Education	Employment	Personality	Looks
0	High School	Unemployed	Introvert	Below Average
1	Undergrad	Blue collar	Extrovert	Average
2	Post grad	White collar		Above Average

Code	High School	Undergrad
0	**1**	**0**
1	**0**	**1**
2	**0**	**0**

Code	Introvert
0	**1**
1	**0**

› In this case, the utility function would look

$$V_j = \alpha_j + \beta_{hs} hs + \beta_{ug} ug + \beta_{um} um + \beta_{bc} bc$$
$$+ \beta_{In} in + \beta_{ba} ba + \beta_{av} av$$

Handling soft (qualitative variables)

Effects coding

› We could use linear coding for this such that

	Person A	Person B	None
Education	Post Grad	High School	
Employment	Unemployed	White collar	
Personality	Extrovert	Introvert	
Looks	Above average	Average	
I would choose			

Levels		
High School	Undergrad	Post grad
Unemployed	Blue collar	White collar
Introvert	Extrovert	
Below Average	Average	Above Average

Code	Education	Employment	Personality	Looks
0	High School	Unemployed	Introvert	Below Average
1	Undergrad	Blue collar	Extrovert	Average
2	Post grad	White collar		Above Average

Code	High School	Undergrad
0	**1**	**0**
1	**0**	**1**
2	**-1**	**-1**

Code	Introvert
0	**1**
1	**-1**

› In this case, the utility function would look

$$V_j = \alpha_j + \beta_{hs} hs + \beta_{ug} ug + \beta_{um} um + \beta_{bc} bc$$
$$+ \beta_{In} in + \beta_{ba} ba + \beta_{av} av$$

Discrete choice modelling process

2

Discrete choice modelling process

Typical modelling steps in discrete choice studies

Experimental design

1	30	3	15	4
2	30	1	35	4
3	30	1	20	4
4	20	1	25	4
5	25	5	30	2
6	20	3	35	2
7	20	1	20	4
8	25	3	40	2
9	25	5	25	2
10	20	5	15	2
11	30	5	30	2
12	25	3	40	4

→

Survey

Please look at each of the following scenarios and choose the alternative you most prefer in each

Task 1 of 3

	Oyster Bay	Casa del Cerro	Fazi Battaglia	None
Cost (per bottle)	€14	€10	€8	
Grape	M	Riesling	Verdicchio	
Year bottled	2009	2008	2009	
Country of Origin	New Zealand	Italy	Chile	
I would choose:	○	○	○	○

Task 2 of 3

	Oyster Bay	Casa del Cerro	Fazi Battaglia	None
Cost (per bottle)	€14	€10	€8	
Grape	M	Riesling	Verdicchio	
Year bottled	2009	2008	2009	
Country of Origin	New Zealand	Italy	Chile	
I would choose:	○	○	○	○

Task 3 of 3

	Oyster Bay	Casa del Cerro	Fazi Battaglia	None
Cost (per bottle)	€14	€10	€8	
Grape	M	Riesling	Verdicchio	
Year bottled	2009	2008	2009	
Country of Origin	New Zealand	Italy	Chile	
I would choose:	○	○	○	○

→

Respondents

→

Data

0 0 1 0
0 1 0 0
0 1 0 0
0 0 0 1
1 0 0 0
0 1 0 0
1 0 0 0

→

Estimation

↓

Specification of:
1. Model type
2. Utility functions

$$V_1 = \sum_{k=1}^{K} \beta_{1k} x_{1k}$$

$$V_2 = \sum_{k=1}^{K} \beta_{2k} x_{2k}$$

$$\vdots$$

$$V_j = \sum_{k=1}^{K} \beta_{jk} x_{jk}$$

←

$$\hat{\beta}$$

$$se(\hat{\beta})$$

Results

Typical modelling steps in discrete choice studies

Experimental design

1	30	3	15	4
2	30	1	35	4
3	30	1	20	4
4	20	1	25	4
5	25	5	30	2
6	20	3	35	2
7	20	1	20	4
8	25	3	40	2
9	25	5	25	2
10	20	5	15	2
11	30	5	30	2
12	25	3	40	4

→

Survey

Please look at each of the following scenarios and choose the alternative you most prefer in each

Task 1 of 3

	Oyster Bay	Casa del Cerro	Fazi Battaglia	None
Cost (per bottle)	€14	€10	€8	
Grape	M	Riesling	Verdicchio	
Year bottled	2009	2008	2009	
Country of Origin	New Zealand	Italy	Chile	
I would choose:	○	○	○	○

Task 2 of 3

	Oyster Bay	Casa del Cerro	Fazi Battaglia	None
Cost (per bottle)	€14	€10	€8	
Grape	M	Riesling	Verdicchio	
Year bottled	2009	2008	2009	
Country of Origin	New Zealand	Italy	Chile	
I would choose:	○	○	○	○

Task 3 of 3

	Oyster Bay	Casa del Cerro	Fazi Battaglia	None
Cost (per bottle)	€14	€10	€8	
Grape	M	Riesling	Verdicchio	
Year bottled	2009	2008	2009	
Country of Origin	New Zealand	Italy	Chile	
I would choose:	○	○	○	○

→

Respondents

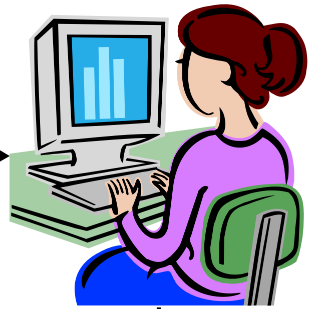
→

Data

0 0 1 0
0 1 0 0
0 1 0 0
0 0 0 1
1 0 0 0
0 1 0 0
1 0 0 0

→

Estimation



Specification of:
1. Model type
2. Utility functions

$$V_1 = \sum_{k=1}^{K} \beta_{1k} x_{1k}$$

$$V_2 = \sum_{k=1}^{K} \beta_{2k} x_{2k}$$

$$\vdots$$

$$V_j = \sum_{k=1}^{K} \beta_{jk} x_{jk}$$

←

$$\hat{\beta}$$

$$se(\hat{\beta})$$

Results

The question is what is the relationship between the results and the experimental design?

Design theory

3

Step 1: Define the lig-likelihood function

› Determine the log-likelihood function:

$$L_N(\beta \mid X, y) = \sum_{n=1}^{N} \sum_{s=1}^{S} \sum_{j=1}^{J} y_{jsn} \log P_{jsn}(X \mid \beta)$$

› Note that:

› In estimation, given the (design) data *X* and the observations *y*, one aims to determine estimates β such that $L_N(\beta \mid X, y)$ is maximised [maximum likelihood estimation]

› When generating an experimental design, these parameter estimates are unknown β

› The values of $P_{jsn}(X \mid \beta)$ depend on the model used (MNL, NL, MMNL)

Step 2: Determine the Fisher Information Matrix

› The second derivatives of the log-likelihood gives the Fisher information matrix:

$$I_N(\beta \mid X, y) = -E_y\left(\frac{\partial^2 L_N(\beta \mid X, y)}{\partial\beta\partial\beta'}\right)$$

[expected negative Hessian matrix of second derivatives]

› The negative inverse of the Fisher information matrix yields the model variance-covariance matrix

$$\Omega_N(\beta \mid X, y) = -I_N^{-1}(\beta \mid X, y)$$

[negative inverse matrix]

MNL model AVC I

› Example: MNL model with generic parameters (McFadden, 1974)

$$L_N(\beta \mid X, y) = \sum_{n=1}^{N} \sum_{s=1}^{S} \sum_{j=1}^{J} y_{jsn} \log P_{jsn}(X \mid \beta)$$

$$P_{jsn}(X \mid \beta) = \frac{\exp\left(V_{jsn}(X \mid \beta)\right)}{\sum_{i=1}^{J} \exp\left(V_{jsn}(X, \mid \beta)\right)} \quad \text{with} \quad V_{jsn}(X \mid \beta) = \sum_{k=1}^{K} \beta_k X_{jksn}$$

First derivative:

$$\frac{\partial L_N(\beta \mid X, y)}{\partial \beta_k} = \sum_{n=1}^{N} \sum_{s=1}^{S} \sum_{j=1}^{J} \left(y_{jsn} - P_{jsn}(X \mid \beta)\right) X_{jksn}$$

Second derivative:

$$\frac{\partial^2 L_N(\beta \mid X)}{\partial \beta_{k_1} \partial \beta_{k_2}} = \sum_{n=1}^{N} \sum_{s=1}^{S} \sum_{j=1}^{J} X_{jk_1 sn} P_{jsn}(X \mid \beta) \left(X_{jk_2 sn} - \sum_{i=1}^{J} X_{ik_2 sn} P_{isn}(X \mid \beta) \right)$$

Note: *y* drops out!

MNL model AVC II

› Example: MNL model with generic parameters (McFadden, 1974)

$$I_N(\beta \mid X) = -\sum_{n=1}^{N}\sum_{s=1}^{S}\sum_{j=1}^{J} X_{jk_1sn} P_{jsn}(X \mid \beta)\left(X_{jk_2sn} - \sum_{i=1}^{J} X_{ik_2sn} P_{isn}(X \mid \beta) \right)$$

Assuming that all responds observe the same choice situations,

$$I_N(\beta \mid X) = -N \cdot \sum_{s=1}^{S}\sum_{j=1}^{J} X_{jk_1s} P_{js}(X \mid \beta)\left(X_{jk_2s} - \sum_{i=1}^{J} X_{ik_2s} P_{is}(X \mid \beta) \right)$$

$$= N \cdot I_1(\beta \mid X)$$

Therefore, the AVC matrix becomes:

$$\Omega_N(\beta \mid X) = I_N^{-1}(\beta \mid X)$$

$$= \tfrac{1}{N} \cdot I_1^{-1}(\beta \mid X)$$

Sample size and efficiency

$$\Omega_N(\beta \mid X) = \tfrac{1}{N} \cdot \Omega_1(\beta \mid X)$$

$$\Rightarrow \quad se_N(\beta \mid X) = \tfrac{1}{\sqrt{N}} \cdot se_1(\beta \mid X)$$

Measuring efficiency

- In order to assess the efficiency of different designs, several efficiency measures have been proposed
- The most widely used ones are:
  - *D*-error $= \det\left(\Omega_1\right)^{1/K}$ [determinant of AVC matrix]
  - *A*-error $= \dfrac{\mathrm{tr}\left(\Omega_1\right)}{K}$ [trace of AVC matrix]

  $K =$ number of parameters (size of the matrix), used as a scaling factor for the efficiency measure
- The lower the *D*-error or *A*-error, the more efficient the design

Orthogonal designs

- The vast majority of designs employed today are orthogonal designs
- An orthogonal design is a design in which the attributes of the design are independent of each other or uncorrelated
- A design will be orthogonal when
  - every pair of levels occurs equally often across all pairs of attributes, the design is said to be orthogonal

S	x_1	x_2
1	0	1
2	0	1
3	0	0
4	0	0
5	0	0
6	0	1
7	1	0
8	1	1
9	1	1
10	1	0
11	1	1
12	1	0

Correl	x_2
x_1	0

  - or when the frequencies for level pairs are proportional instead of equal.

S	x_1	x_2
1	0	1
2	0	1
3	0	0
4	0	0
5	1	0
6	1	1
7	1	0
8	1	1
9	1	1
10	1	0
11	1	1
12	1	0

Correl	x_2
x_1	0

But how is orthogonality constructed?

› Orthogonality is defined and constructed in relation to the design codes

S	a	b	c	d	e	f
1	-1	1	1	1	0	-1
2	1	1	-1	0	1	0
3	-1	1	1	-1	0	1
4	1	0	0	-1	-1	-1
5	0	0	0	-1	-1	-1
6	1	-1	1	1	-1	1
7	0	1	-1	1	-1	1
8	0	-1	1	0	1	0
9	0	0	0	0	1	0
10	-1	-1	-1	1	0	-1
11	1	0	0	0	1	0
12	-1	-1	-1	-1	0	1

→

	a	b	c	d	e	f
a	**1**	0	0	0	0	0
b	0	**1**	**0**	**0**	**0**	**0**
c	0	0	**1**	**0**	**0**	**0**
d	0	0	0	**1**	**0**	**0**
e	0	0	0	0	**1**	**0**
f	0	0	0	0	0	**1**

But what about dummy coding?

› Orthogonality is defined and constructed in relation to the design codes

S	a	b	c	d	e	f
1	-1	1	1	1	0	-1
2	1	1	-1	0	1	0
3	-1	1	1	-1	0	1
4	1	0	0	-1	-1	-1
5	0	0	0	-1	-1	-1
6	1	-1	1	1	-1	1
7	0	1	-1	1	-1	1
8	0	-1	1	0	1	0
9	0	0	0	0	1	0
10	-1	-1	-1	1	0	-1
11	1	0	0	0	1	0
12	-1	-1	-1	-1	0	1

S	a1	a2	b1	b2	c1	c2	d1	d2	e1	e2	f1	f2
1	1	0	0	0	0	0	0	0	0	1	1	0
2	0	0	0	0	1	0	0	1	0	0	0	1
3	1	0	0	0	0	0	1	0	0	1	0	0
4	0	0	0	1	0	1	1	0	1	0	1	0
5	0	1	0	1	0	1	1	0	1	0	1	0
6	0	0	1	0	0	0	0	0	1	0	0	0
7	0	1	0	0	1	0	0					
8	0	1	1	0	0	0						
9	0	1	0	1	0	1						
10	1	0	1	0	1	0						
11	0	0	0	1	0	1						
12	1	0	1	0	1	0						

	a1	a2	b1	b2	c1	c2	d1	d2	e1	e2	f1	f2
a1	**1**	-0.5	0.25	-0.5	0.25	-0.5	0.25	-0.5	-0.5	1	0.25	-0.5
a2	-0.5	**1**	-0.125	0.25	-0.125	0.25	-0.125	0.25	0.25	-0.5	-0.125	0.25
b1	0.25	-0.125	**1**	-0.5	0.25	-0.5	-0.125	-0.125	-0.125	0.25	-0.125	-0.125
b2	-0.5	0.25	-0.5	**1**	-0.5	1	0.25	0.25	0.25	-0.5	0.25	0.25
c1	0.25	-0.125	0.25	-0.5	**1**	-0.5	-0.125	-0.125	-0.125	0.25	-0.125	-0.125
c2	-0.5	0.25	-0.5	1	-0.5	**1**	0.25	0.25	0.25	-0.5	0.25	0.25
d1	0.25	-0.125	-0.125	0.25	-0.125	0.25	**1**	-0.5	0.25	0.25	0.25	-0.5
d2	-0.5	0.25	-0.125	0.25	-0.125	0.25	-0.5	**1**	-0.5	-0.5	-0.5	1
e1	-0.5	0.25	-0.125	0.25	-0.125	0.25	0.25	-0.5	**1**	-0.5	0.25	-0.5
e2	1	-0.5	0.25	-0.5	0.25	-0.5	0.25	-0.5	-0.5	**1**	0.25	-0.5
f1	0.25	-0.125	-0.125	0.25	-0.125	0.25	0.25	-0.5	0.25	0.25	**1**	-0.5
f2	-0.5	0.25	-0.125	0.25	-0.125	0.25	-0.5	1	-0.5	-0.5	-0.5	**1**

But what about effects coding?

› Orthogonality is defined and constructed in relation to the design codes

S	a	b	c	d	e	f
1	-1	1	1	1	0	-1
2	1	1	-1	0	1	0
3	-1	1	1	-1	0	1
4	1	0	0	-1	-1	-1
5	0	0	0	-1	-1	-1
6	1	-1	1	1	-1	1
7	0	1	-1	1	-1	1
8	0	-1	1	0	1	0
9	0	0	0	0	1	0
10	-1	-1	-1	1	0	-1
11	1	0	0	0	1	0
12	-1	-1	-1	-1	0	1

S	a1	a2	b1	b2	c1	c2	d1	d2	e1	e2	f1	f2
1	1	0	-1	-1	-1	-1	-1	-1	0	1	1	0
2	-1	-1	-1	-1	1	0	0	1	-1	-1	0	1
3	1	0	-1	-1	-1	-1	1	0	0	1	-1	-1
4	-1	-1	0	1	0	1	1	0	1	0	1	0
5	0	1	0	1	0	1	1	0	1	0	1	0
6	-1	-1	1	0	-1	-1	-1	-1	1	0	-1	-1
7	0	1	-1	-1	1	0	-1					
8	0	1	1	0	-1	-1						
9	0	1	0	1	0	1						
10	1	0	1	0	1	0						
11	-1	-1	0	1	0	1						
12	1	0	1	0	1	0						

	a1	a2	b1	b2	c1	c2	d1	d2	e1	e2	f1	f2
a1	**1**	0.5	0	-0.4	0	-0.4	0	-0.375	0	0.75	0	-0.375
a2	0.5	**1**	0	0	0	0	0	0	0	0	0	0
b1	0	0	**1**	0.5	0	0	0	0	0	0	0	0
b2	-0.4	0	0.5	**1**	0	0.75	0.375	0.375	0	-0.4	0.375	0.375
c1	0	0	0	0	**1**	0.5	0	0	0	0	0	0
c2	-0.4	0	0	0.75	0.5	**1**	0.375	0.375	0	-0.4	0.375	0.375
d1	0	0	0	0.38	0	0.38	**1**	0.5	0	0	0	0
d2	-0.4	0	0	0.38	0	0.38	0.5	**1**	-0.75	-0.8	0	0.75
e1	0	0	0	0	0	0	0	-0.75	**1**	0.5	0	-0.75
e2	0.75	0	0	-0.4	0	-0.4	0	-0.75	0.5	**1**	0	-0.75
f1	0	0	0	0.38	0	0.38	0	0	0	0	**1**	0.5
f2	-0.4	0	0	0.38	0	0.38	0	0.75	-0.75	-0.8	0.5	**1**

Or orthogonal coding

› Orthogonality is defined and constructed in relation to the design codes

S	a	b	c	d	e	f
1	-1	1	1	1	0	-1
2	1	1	-1	0	1	0
3	-1	1	1	-1	0	1
4	1	0	0	-1	-1	-1
5	0	0	0	-1	-1	-1
6	1	-1	1	1	-1	1
7	0	1	-1	1	-1	1
8	0	-1	1	0	1	0
9	0	0	0	0	1	0
10	-1	-1	-1	1	0	-1
11	1	0	0	0	1	0
12	-1	-1	-1	-1	0	1

S	a1	a2	b1	b2	c1	c2	d1	d2	e1	e2	f1	f2
1	-1	1	1	1	1	1	1	1	0	-2	-1	1
2	1	1	1	1	-1	1	0	-2	1	1	0	-2
3	-1	1	1	1	1	1	-1	1	0	-2	1	1
4	1	1	0	-2	0	-2	-1	1	-1	1	-1	1
5	0	-2	0	-2	0	-2	-1	1	-1	1	-1	1
6	1	1	-1	1	1	1	1	1	-1	1	1	1
7	0	-2	1	1	-1	1						
8	0	-2	-1	1	1							
9	0	-2	0	-2	0							
10	-1	1	-1	1	-1							
11	1	1	0	-2	0							
12	-1	1	-1	1	-1							

	a1	a2	b1	b2	c1	c2	d1	d2	e1	e2	f1	f2
a1	**1**	0	0	-0.433	0	-0.433	0	-0.433	0	0.866	0	-0.433
a2	0	**1**	0	0.25	0	0.25	0	0.25	0	-0.5	0	0.25
b1	0	0	**1**	0	0	0	0	0	0	0	0	0
b2	-0.43	0.3	0	**1**	0	1	0.433	0.25	0	-0.5	0.433	0.25
c1	0	0	0	0	**1**	0	0	0	0	0	0	0
c2	-0.43	0.3	0	1	0	**1**	0.433	0.25	0	-0.5	0.433	0.25
d1	0	0	0	0.433	0	0.433	**1**	0	0	0	0	0
d2	-0.43	0.3	0	0.25	0	0.25	0	**1**	-0.87	-0.5	0	1
e1	0	0	0	0	0	0	0	-0.866	**1**	0	0	-0.866
e2	0.866	-1	0	-0.5	0	-0.5	0	-0.5	0	**1**	0	-0.5
f1	0	0	0	0.433	0	0.433	0	0	0	0	**1**	0
f2	-0.43	0.3	0	0.25	0	0.25	0	1	-0.87	-0.5	0	**1**

Qualitative variables and experimental design

4

THE UNIVERSITY OF SYDNEY

Issue #1: Degrees of freedom...

- **The statistical design perspective**
- Each parameter requires a degree of freedom (d.f.) to be estimated
  - This includes, ASCs, main effects, interaction effects, generic parameters, alternative specific parameters, parameters aligned with non-linear marginal utility effects (dummy, effects, orthogonal codes), standard deviation parameters, etc.

> ! This is why we have been writing out the expected utility functions and deciding what form each function might take

- The estimated model will estimate *K* parameters

Issue #1: Theory

- **The statistical design perspective**
- Each choice task contains up to J alternatives and up to J-1 uniquely identified and independent observations
  - If we can calculate the choice probabilities of the first J-1 alternatives in a choice task, we can determine the probability for the last J^{th} alternative
  - As such, each choice task, S, contains up to J-1 independent choice observations (pieces of information) for modelling purposes

!

Overall, the design will have $S\times(J\text{-}1)$ independent choice observations

Issue #1: More parameters, larger design

- **The statistical design perspective**
- The total number of independent choice probabilities should at least be equal to or greater than the number of parameters to be estimated

!

$$S \times (J\text{-}1) > K$$

$$V_j = \alpha_j + \beta_{ed} ed + \beta_{Em} em + \beta_{pe} pe + \beta_{lo} lo$$

$$V_j = \alpha_j + \beta_{hs} hs + \beta_{ug} ug + \beta_{um} um + \beta_{bc} bc + \beta_{In} in + \beta_{ba} ba + \beta_{av} av$$

Issue #2: Sparse matrix I

› The design matrix for a design with dummy/effects codes can be quite sparse (lots of zeros)

	Alt. A				Alt. B			
S	Edu.	Emp.	Per.	Lks	Edu.	Emp.	Per.	Lks
1	1	2	1	2	2	1	0	1
2	1	2	0	1	2	1	1	2
3	2	2	1	1	0	0	0	2
4	2	1	1	1	1	0	0	2
5	0	1	0	2	2	2	1	0
6	1	0	0	2	2	2	1	1
7	1	2	0	0	0	1	1	2
8	0	0	1	1	0	1	0	0
9	2	1	0	0	0	2	1	2
10	1	1	1	2	2	0	0	1
11	0	0	1	0	1	0	1	0
12	2	0	1	2	0	2	0	1
13	0	1	0	0	2	0	1	1
14	0	2	1	2	1	1	0	0
15	2	0	0	0	1	1	0	1
16	1	1	1	1	0	2	0	0
17	2	2	0	0	1	0	1	2
18	0	0	0	1	1	2	1	0

	Alt. A							Alt. B						
S	Edu. HS	Edu. UG	Emp. UE	Emp. BC	Per. Int	Lks. BA	Lks AV	Edu. HS	Edu. UG	Emp. UE	Emp. BC	Per. Int	Lks. BA	Lks AV
1	0	1	0	0	1	0	0	0	0	0	1	0	0	1
2	0	1	0	0	0	0	1	0	0	0	1	1	0	0
3	0	0	0	0	1	0	1	1	0	1	0	0	0	0
4	0	0	0	1	1	0	1	0	1	1	0	0	0	0
5	1	0	0	1	0	0	0	0	0	0	0	1	1	0
6	0	1	1	0	0	0	0	0	0	0	0	1	0	1
7	0	1	0	0	0	1	0	1	0	0	1	1	0	0
8	1	0	1	0	1	0	1	1	0	0	1	0	1	0
9	0	0	0	1	0	1	0	1	0	0	0	1	0	0
10	0	1	0	1	1	0	0	0	0	1	0	0	0	1
11	1	0	1	0	1	1	0	0	1	1	0	1	1	0
12	0	0	1	0	1	0	0	1	0	0	0	0	0	1
13	1	0	0	1	0	1	0	0	0	1	0	1	0	1
14	1	0	0	0	1	0	0	0	1	0	1	0	1	0
15	0	0	1	0	0	1	0	0	1	0	1	0	0	1
16	0	1	0	1	1	0	1	1	0	0	0	0	1	0
17	0	0	0	0	0	1	0	0	1	1	0	1	0	0
18	1	0	1	0	0	0	1	0	1	0	0	1	1	0

Issue #2: Sparse matrix II

› The problem increases as the number of attribute levels increases

	Alt. A				Alt. B			
S	Edu.	Emp.	Per.	Lks	Edu.	Emp.	Per.	Lks
1	2	1	1	0	3	0	1	1
2	5	1	1	0	1	0	0	1
3	5	2	1	2	4	1	0	2
4	2	0	0	2	5	2	1	1
5	1	0	1	2	4	1	1	0
6	1	2	1	1	5	0	0	0
7	3	0	0	1	0	2	1	2
8	3	1	0	1	2	2	0	0
9	5	1	0	1	3	2	1	2
10	4	1	1	0	0	0	1	1
11	1	0	0	0	0	1	0	0
12	0	0	0	2	2	2	1	1
13	2	0	1	2	1	2	0	0
14	3	2	0	0	4	0	1	2
15	4	2	1	0	2	1	0	1
16	0	2	0	1	1	1	0	2
17	0	1	1	2	5	0	0	0
18	4	2	0	1	3	1	1	2

	Alt. A										Alt. B									
S	Edu 1	Edu 2	Edu 3	Edu 4	Edu 5	Emp. UE	Emp. BC	Per. Int	Lks. BA	Lks AV	Edu 1	Edu 2	Edu 3	Edu 4	Edu 5	Emp. UE	Emp. BC	Per. Int	Lks. BA	Lks AV
1	0	0	1	0	0	0	1	1	1	0	0	0	0	1	0	1	0	1	0	1
2	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	1	0	0	0	1
3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0
4	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1
5	0	1	0	0	0	1	0	1	0	0	0	0	0	0	1	0	1	1	1	0
6	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	1	0
7	0	0	0	1	0	1	0	0	0	1	1	0	0	0	0	0	0	1	0	0
8	0	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1	0
9	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0
10	0	0	0	0	1	0	1	1	1	0	1	0	0	0	0	1	0	1	0	1
11	0	1	0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	1	0
12	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	1
13	0	0	1	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	1	0
14	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	1	0	1	0	0
15	0	0	0	0	1	0	0	1	1	0	0	0	1	0	0	0	1	0	0	1
16	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0
17	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	1	0
18	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	1	0	0

Issue #2: Sparse matrix III

› When the design matrix *X* is sparse, you have fewer data points per column

S	Edu.
1	2
2	5
3	5
4	2
5	1
6	1
7	3
8	3
9	5
10	4
11	1
12	0
13	2
14	3
15	4
16	0
17	0
18	4

S	Edu 1	Edu2	Edu3	Edu 4	Edu 5
1	0	0	1	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	1	0	0
5	0	1	0	0	0
6	0	1	0	0	0
7	0	0	0	1	0
8	0	0	0	1	0
9	0	0	0	0	0
10	0	0	0	0	1
11	0	1	0	0	0
12	1	0	0	0	0
13	0	0	1	0	0
14	0	0	0	1	0
15	0	0	0	0	1
16	1	0	0	0	0
17	1	0	0	0	0
18	0	0	0	0	1

› This can be problem when calculating the Fisher matrix

› It can be a real problem when taking the inverse of the Fisher matrix

$$\Omega_N(\beta \mid X, y) = -I_N^{-1}(\beta \mid X, y)$$

Issue #2: Sparse matrix IV

› When the design matrix *X* is sparse, you have fewer data points per column

S	Edu.
1	2
2	5
3	5
4	2
5	1
6	1
7	3
8	3
9	5
10	4
11	1
12	0
13	2
14	3
15	4
16	0
17	0
18	4

S	Edu 1	Edu2	Edu3	Edu 4	Edu 5
1	0	0	1	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	1	0	0
5	0	1	0	0	0
6	0	1	0	0	0
7	0	0	0	1	0
8	0	0	0	1	0
9	0	0	0	0	0
10	0	0	0	0	1
11	0	1	0	0	0
12	1	0	0	0	0
13	0	0	1	0	0
14	0	0	0	1	0
15	0	0	0	0	1
16	1	0	0	0	0
17	1	0	0	0	0
18	0	0	0	0	1

› **Solutions**:

1. Increase the number of rows in the design
2. Assume more than one respondent when generating the design
3. Decrease the number of levels assumed

Issue #3: Finding priors I

- The preponderance of quantitative variables makes finding priors for these types of attributes generally much easier
  - How many studies that have used travel time and cost attributes?
  - What prior should you use for seat colour though?
- The direction of the priors for quantitative variables is generally easier to guess *a priori*
  - Cost and travel time parameters should be –ve
  - What is preferred seat colour: black, green, or white or red?

Issue #3: Finding priors II

› Typically we would do a pilot study to get priors

› Likely to be more heterogeneity surrounding qualitative variables however

- Most people would prefer less cost than more in the population
- Seat colour preference is likely to be less homogenous over the population

› **Solution**: use Bayesian priors

- Instead of assuming exact knowledge about the priors, i.e., β_k, assume a possible distribution for the priors $\beta_k \sim U(\mu_L, \mu_U)$
- If you assume uniform priors, the bounds can be either side of zero, hence not assume specific direction about preference

$$\int_{\beta} \det\left(\Omega(\beta \mid X)\right)^{1/K} f(\beta \mid \mu_L, \mu_U) d\beta$$

Uniform probability density function

Issue #3: Finding priors III

› Solution: use Bayesian priors

- Instead of assuming exact knowledge about the priors, i.e., β_k, assume a possible distribution for the priors $\beta_k \sim U(\mu_L, \mu_U)$
- If you assume uniform priors, the bounds can be either side of zero, hence not assume specific direction about preference

› **Problem**: If you make the priors to wide, then the higher the probability of obtaining outliers in your efficiency measure increase

$$\int_\beta \det\left(\Omega(\beta \mid X)\right)^{1/K} f(\beta \mid \mu_L, \mu_U) d\beta$$

- This may make the design generation process unstable, or produce not very good designs, as outlier draws may dominate the measure

Benefit 1: Attribute level balance doesn't matter I

› Depending on the priors assumed, the most statistically efficient design will tend to be an end-point design

- That is, one that assumes the widest levels only

U(A) = ed[-0.5]*ed[0,1,2] + em[-0.5]*em[0,1,2] + pe[-0.5]*pe[1,0] + lo[-0.75]*lo[0,1,2]
U(B) = ed[-0.5]*ed[0,1,2] + em[-0.5]*em[0,1,2] + pe[-0.5]*pe[1,0] + lo[-0.75]*lo[0,1,2]
U(C) = 0

	Alt. A				Alt. B			
S	Edu.	Emp.	Per.	Lks	Edu.	Emp.	Per.	Lks
1	0	2	1	0	0	0	0	2
2	2	0	0	0	0	0	1	2
3	2	2	0	0	0	0	1	2
4	2	0	0	0	0	2	1	0
5	2	2	0	0	0	0	1	2
6	0	0	1	0	0	2	0	0
7	0	2	0	0	2	0	1	0
8	0	0	0	2	1	0	1	0
9	2	0	0	1	0	2	1	0
10	2	0	1	0	0	2	0	1
11	2	0	1	0	0	0	0	2
12	0	1	0	2	2	1	0	0

S	Edu. HS	Edu. UG
1	1	**0**
2	0	**0**
3	0	**0**
4	0	**0**
5	0	**0**
6	1	**0**
7	1	**0**
8	1	**0**
9	0	**0**
10	0	**0**
11	0	**0**
12	1	**0**

Benefit 1: Attribute level balance doesn't matter II

› If you generate the design specifically for dummy/effects codes, the previous problem cannot occur, as a design column of all zeros will led to a singular AVC matrix, and hence the design will be discarded automatically

U(A) = ed.dummy[-0.5|-0.1]*ed + em.dummy[-0.5|-0.1]*em + pe.dummy[-0.5]*pe[1,0] + lo.dummy[-0.75|-0.2]*lo
U(B) = ed.dummy[-0.5|-0.1]*ed + em.dummy[-0.5|-0.1]*em + pe.dummy[-0.5]*pe[1,0] + lo.dummy[-0.75|-0.2]*lo
$ U(C) = 0

	Alt. A				Alt. B			
S	Edu.	Emp.	Per.	Lks	Edu.	Emp.	Per.	Lks
1	0	1	0	1	1	0	1	0
2	2	0	1	1	0	1	0	2
3	0	2	0	0	1	1	1	2
4	0	2	0	1	2	1	1	0
5	1	2	1	0	0	0	0	2
6	2	0	0	1	1	2	1	2
7	2	1	1	1	0	2	0	0
8	0	0	1	2	1	1	0	1
9	1	1	0	0	0	0	1	2
10	1	2	1	1	2	1	0	0
11	1	0	0	2	0	2	1	1
12	0	1	1	2	1	0	0	1

S	Edu. HS	Edu. UG
1	1	0
2	0	0
3	1	0
4	1	0
5	0	1
6	0	0
7	0	0
8	1	0
9	0	1
10	0	1
11	0	1
12	1	0

Conclusions about designs for qualitative variables

5

Conclusions about designs for qualitative variables

Summary

- Harder than dealing with quantitative attributes
  - Harder to get priors
  - Will typically require larger designs due to
    - Sparse X matrix
    - More degrees of freedom
- But better able to handle non-balance

One more thing

6

Cooper, B, Rose, J.M. and Crase, L. (2012) Does anybody like water restrictions? Some observations in Australian urban communities, *The Australian Journal of Agricultural and Resource Economics*, 56(1), 61-81

THE UNIVERSITY OF SYDNEY

A common mistake I

- It is not uncommon for researchers to construct SP experiments with a status quo alternative that has attribute levels
- When doing this, it is common to set to assign the base level of qualitative attributes to the SQ alternative only

Which enforcement & education		Price of the enforcement package	Number of inspectors	Information	Able to report your neighbour
			…00 …lds	Every 14 days	Yes
			…00 households	Every 7 days	No
Neither		$0 per year	1 per 10 000 households	Every 90 days	No

Information attribute levels	Original code	Dummy code			
		Everyday	7 days	14 days	31 days
Everyday	0	1	0	0	0
Every 7 days	1	0	1	0	0
Every 14 days	2	0	0	1	0
Every 31 days	3	0	0	0	1
Every 90 days	4	0	0	0	0

A common mistake II

- Now the base level is perfectly confounded with the SQ alternative
  - This is theoretically problem – these are dummy/effects codes at the attributes level, not alternative level
- This creates problems with estimating ASCs as the base level is now perfectly confounded with an alternative

	Alt. A				Alt. B				SQ			
S	Edu.	Emp.	Per.	Lks	Edu.	Emp.	Per.	Lks	Edu.	Emp.	Per.	Lks
1	0	1	0	1	1	0	1	0	3	3	2	3
2	2	0	1	1	0	1	0	2	3	3	2	3
3	0	2	0	0	1	1	1	2	3	3	2	3
4	0	2	0	1	2	1	1	0	3	3	2	3
5	1	2	1	0	0	0	0	2	3	3	2	3
6	2	0	0	1	1	2	1	2	3	3	2	3
7	2	1	1	1	0	2	0	0	3	3	2	3
8	0	0	1	2	1	1	0	1	3	3	2	3
9	1	1	0	0	0	0	1	2	3	3	2	3
10	1	2	1	1	2	1	0	0	3	3	2	3
11	1	0	0	2	0	2	1	1	3	3	2	3
12	0	1	1	2	1	0	0	1	3	3	2	3

A common mistake III

› This requires the use of a hybrid dummy/effects coding where you have two base levels

Information attribute levels	Original code	Hybrid code		
		Everyday	7 days	14 days
Everyday	0	1	0	0
Every 7 days	1	0	1	0
Every 14 days	2	0	0	1
Every 31 days	3	−1	−1	−1
Every 90 days	4	0	0	0