Efficient designs

Key concepts & study plan

Experimental design

Data collection & processing

Model specification & estimation

Interpretation & application

Efficient designs

Outline

- Determining design efficiency
- Generating efficient designs
- Sample size requirements
- Relevant videos to watch

Determining design efficiency

Key concepts & study plan

Experimental design

Data collection & processing

Model specification & estimation

Interpretation & application

Determining design efficiency

Efficiency

- An experimental design is more efficient if it captures more information
- More information means more precise parameter estimates

Determining design efficiency

Remember …

- Covariance matrix is output of model estimation
- Standard errors are derived from the covariance matrix

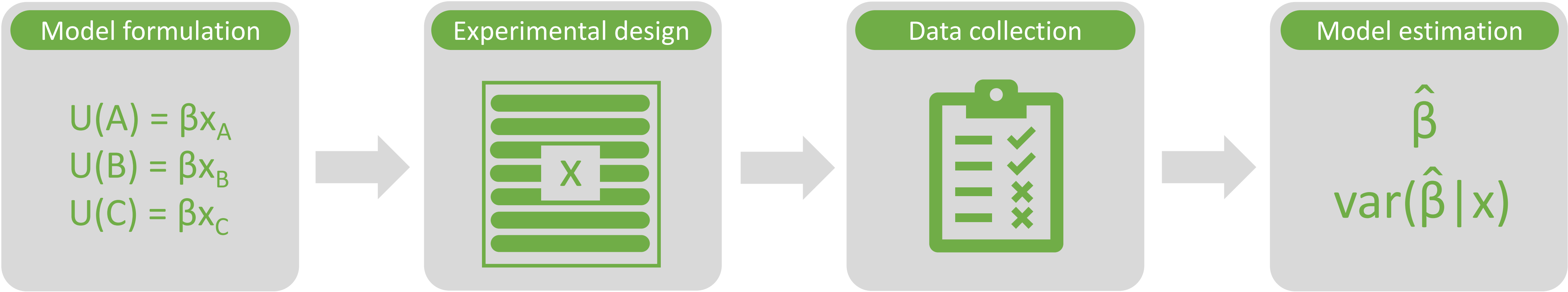
$$\text{var}(\hat{\boldsymbol{\beta}} \mid \mathbf{x}) = \begin{pmatrix} 0.36 & -0.14 & -0.07 \\ -0.14 & 0.79 & 0.06 \\ -0.07 & 0.06 & 0.12 \end{pmatrix}$$

$$se(\hat{\beta}_1 \mid \mathbf{x}) = \sqrt{0.36} = 0.60$$
$$se(\hat{\beta}_2 \mid \mathbf{x}) = \sqrt{0.79} = 0.89$$
$$se(\hat{\beta}_3 \mid \mathbf{x}) = \sqrt{0.12} = 0.35$$

Determining design efficiency

Typical process for choice modelling with stated preference data

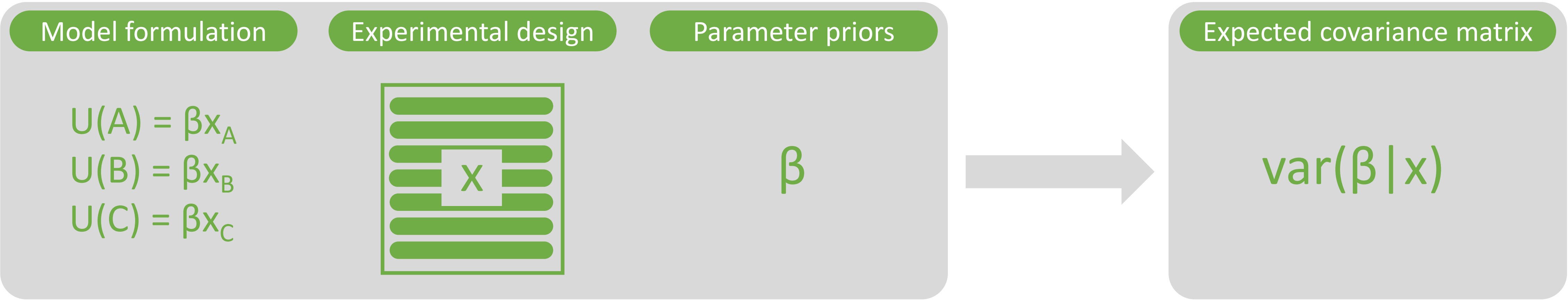
- Experimental design influences the covariance matrix (and thus standard errors)



Determining design efficiency

Predicting covariance matrix *before* data collection

- Expected covariance matrix does not depend on choice observations (**y**)
- Expected covariance matrix can be determined a priori based on parameter priors



Determining design efficiency

Expected covariance matrix for MNL model

- $\text{var}(\boldsymbol{\beta}\,|\,\mathbf{x}) = \left[I(\boldsymbol{\beta}\,|\,\mathbf{x})\right]^{-1}$

  where $I(\boldsymbol{\beta}\,|\,\mathbf{x}) = \sum_n \sum_t \sum_j (\mathbf{x}_{ntj} - \overline{\mathbf{x}}_{nt})' p_{ntj} (\mathbf{x}_{ntj} - \overline{\mathbf{x}}_{nt}),\qquad \overline{\mathbf{x}}_{nt} = \sum_j \mathbf{x}_{ntj} p_{ntj},\qquad p_{ntj} = \frac{\exp(\boldsymbol{\beta}\mathbf{x}'_{ntj})}{\sum_i \exp(\boldsymbol{\beta}\mathbf{x}'_{nti})}$

- $I(\boldsymbol{\beta}\,|\,\mathbf{x})$ is the (Fisher) information matrix
- $\mathbf{x}_{ntj}$ is a row vector of attribute levels for alternative *j* in choice task *t* for decision-maker *n*
- p_{ntj} is the probability that decision-maker *n* chooses alternative *j* in choice task *t*
- $\boldsymbol{\beta}$ is a row vector of parameter values

McFadden, D. (1973) Conditional logit analysis of qualitative choice behavior. In: Zarembka (ed.) *Frontiers in Econometrics*, Academic Press, NY, 105-142.

Determining design efficiency

Expected covariance matrix for MNL model in homogenous choice experiment

- $\text{var}(\boldsymbol{\beta}\,|\,\mathbf{x}) = \left[I(\boldsymbol{\beta}\,|\,\mathbf{x})\right]^{-1}$

  where $I(\boldsymbol{\beta}\,|\,\mathbf{x}) = \mathbf{Z}'\mathbf{Z}, \quad \mathbf{Z} = [\mathbf{z}_{tj}] \in R^{TJ \times K}, \quad \mathbf{z}_{tj} = \left(\mathbf{x}_{tj} - \overline{\mathbf{x}}_t\right)\sqrt{p_{tj}}, \quad \overline{\mathbf{x}}_t = \sum_j \mathbf{x}_{tj} p_{tj}, \quad p_{ntj} = \frac{\exp(\boldsymbol{\beta}\mathbf{x}'_{ntj})}{\sum_i \exp(\boldsymbol{\beta}\mathbf{x}'_{nti})}$

- $I(\boldsymbol{\beta}\,|\,\mathbf{x})$ is the (Fisher) information matrix assuming a single respondent facing all choice tasks
- $\mathbf{x}_{tj}$ is a row vector of attribute levels for alternative *j* in choice task *t*
- p_{tj} is the probability that a decision-maker chooses alternative *j* in choice task *t*
- $\boldsymbol{\beta}$ is a row vector of parameter priors

⚠ Experimental design software, such as Ngene, calculates this for you

Determining design efficiency

Parameter priors

- Best guesses for parameter values
- Informative priors
  - Examples: β = 0.5, β = -0.8
  - Based on pilot study
  - Based on literature study
  - Based on expert judgement
- Uninformative priors
  - Examples: β = 0, β = -0.0001
  - No information, or only information on sign of parameter

Parameter priors

β

The closer the priors are to the actual parameter values, the more efficient the data collection from the choice experiment will be

Determining design efficiency

Different designs result in different covariance matrices

Model formulation

$U(A) = \beta x_A$
$U(B) = \beta x_B$
$U(C) = \beta x_C$

Experimental design

Parameter priors

β

→

Expected covariance matrix

$var(\beta|x)$

Model formulation

$U(A) = \beta x_A$
$U(B) = \beta x_B$
$U(C) = \beta x_C$

Experimental design

Parameter priors

β

→

Expected covariance matrix

$var(\beta|x)$

Determining design efficiency

Example

- For a given choice model and parameter priors, which experimental design is more efficient?

- 

$$\mathrm{var}(\boldsymbol{\beta} \mid \mathbf{x}) = \begin{pmatrix} 0.36 & -0.14 & -0.07 \\ -0.14 & 0.79 & 0.06 \\ -0.07 & 0.06 & 0.12 \end{pmatrix}$$

- 

$$\mathrm{var}(\boldsymbol{\beta} \mid \mathbf{x}) = \begin{pmatrix} 0.51 & -0.39 & -0.13 \\ -0.39 & 0.55 & 0.02 \\ -0.13 & 0.02 & 0.17 \end{pmatrix}$$

Determining design efficiency

Efficiency measures

- D-error
  - Determinant of the expected covariance matrix
  - Lower values indicate more efficient designs

$$\text{D-error} = \left(\det\left(\text{var}(\boldsymbol{\beta} \mid \mathbf{x})\right)\right)^{1/K}$$

- A-error
  - Trace of the expected covariance matrix
  - Lower values indicate more efficient designs

$$\text{A-error} = \frac{\text{tr}\left(\text{var}(\boldsymbol{\beta} \mid \mathbf{x})\right)}{K}$$

$K =$ number of coefficients

Determining design efficiency

Example

- Which experimental design is more efficient?

- 

$$\text{var}(\boldsymbol{\beta} \mid \mathbf{x}) = \begin{pmatrix} 0.36 & -0.14 & -0.07 \\ -0.14 & 0.79 & 0.06 \\ -0.07 & 0.06 & 0.12 \end{pmatrix}$$

$$\text{D-error} = 0.3029$$
$$\text{A-error} = 0.4233$$

- 

$$\text{var}(\boldsymbol{\beta} \mid \mathbf{x}) = \begin{pmatrix} 0.51 & -0.39 & -0.13 \\ -0.39 & 0.55 & 0.02 \\ -0.13 & 0.02 & 0.17 \end{pmatrix}$$

$$\text{D-error} = 0.2430$$
$$\text{A-error} = 0.4100$$

Determining design efficiency

Example

- Which experimental design is more efficient?
- 

$$\mathrm{var}(\boldsymbol{\beta}\,|\,\mathbf{x}) = \begin{pmatrix} 0.36 & -0.14 & -0.07 \\ -0.14 & 0.79 & 0.06 \\ -0.07 & 0.06 & 0.12 \end{pmatrix}$$

D-error $= 0.3029$

A-error $= 0.4233$

E1 =MDETERM(A1:C3)^(1/3)

	A	B	C	D	E
1	0.36	-0.14	-0.07		0.30287968
2	-0.14	0.79	0.06		
3	-0.07	0.06	0.12		

Determining design efficiency

Example

- Which experimental design is more efficient?

- 

$$\operatorname{var}(\boldsymbol{\beta} \mid \mathbf{x}) = \begin{pmatrix} 0.36 & -0.14 & -0.07 \\ -0.14 & 0.79 & 0.06 \\ -0.07 & 0.06 & 0.12 \end{pmatrix}$$

D-error $= 0.3029$

A-error $= 0.4233$

E1 | =(A1+B2+C3)/3

	A	B	C	D	E
1	0.36	-0.14	-0.07		0.42333333
2	-0.14	0.79	0.06		
3	-0.07	0.06	0.12		

Generating efficient designs

Key concepts
& study plan

**Experimental
design**

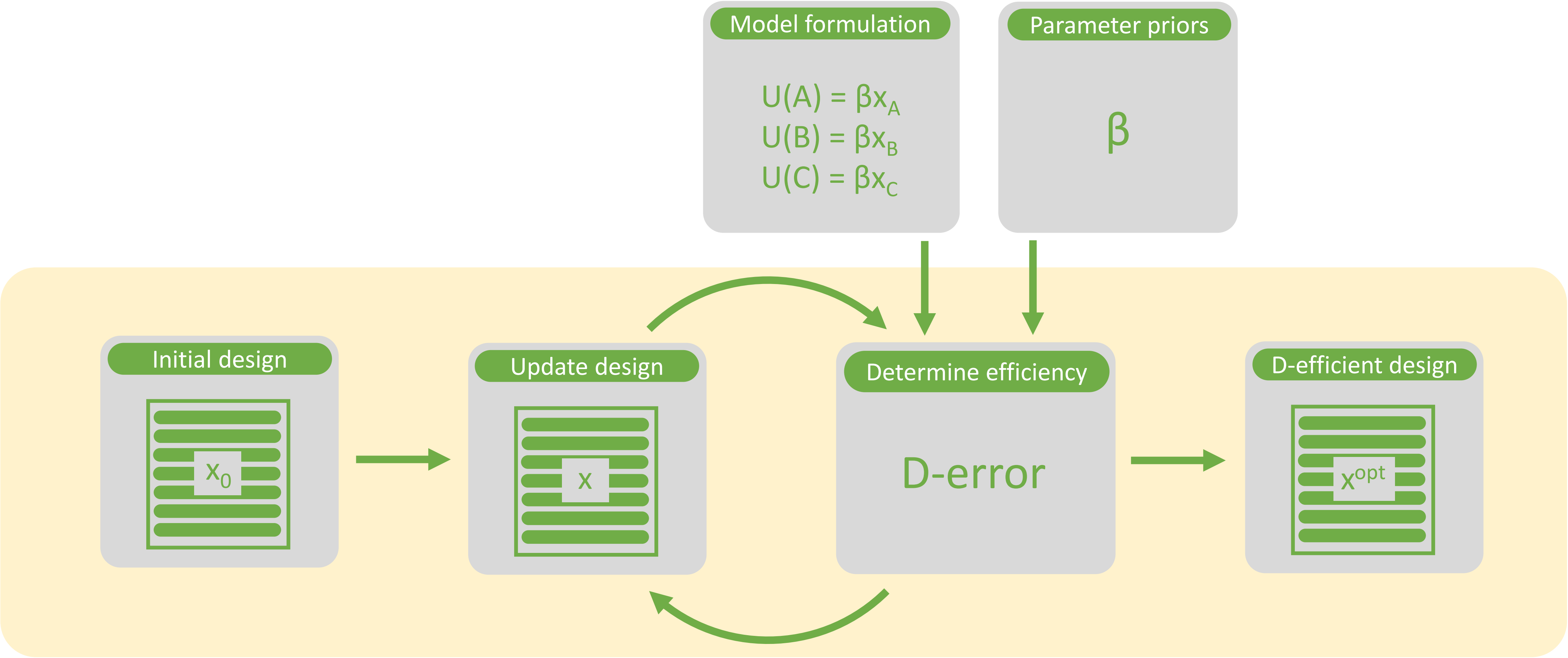
Data collection
& processing

Model specification
& estimation

Interpretation
& application

Generating efficient designs

General algorithm



Generating efficient designs

Swapping algorithm

- Updates design by modifying columns
  - Swapping levels within attributes
- Suitable for designs with column-based constraints
  - Attribute level balance

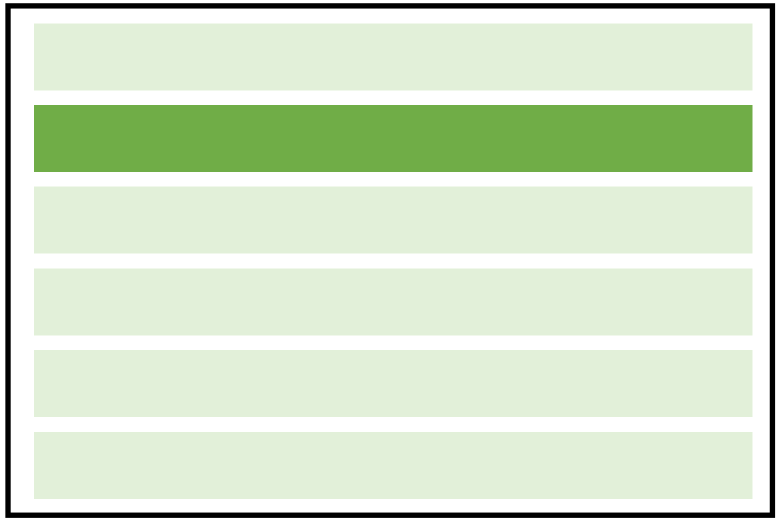
This is the default algorithm in Ngene

Generating efficient designs

Modified Fedorov algorithm

- Updates design by modifying rows
  - Swapping choice tasks from a candidate set
- Suitable for designs with row-based constraints
  - Dominance avoidance
  - Realistic profiles
  - Attribute level overlap

This is an alternative algorithm in Ngene



Generating efficient designs

Example

- See Efficient design – manual generation.xlsx

Utility functions

V1 = b1 + b3*A + b4*B + b5*C
V2 = b2 + b3*A + b4*B + b6*D
V3 = b3*A + b7*E

Parameter priors (beta)

b1	b2	b3	b4	b5	b6	b7
0.5	0.3	-0.2	-0.4	0.5	0.8	-1.0

D-error

0.9445

Experimental design (x) in long format / **Computing logit probabilities**

t	j	asc1	asc2	A	B	C	D	E	V_tj	exp(V_tj)	sum	P_tj
1	1	1	0	10	8	1	0	0	-4.2	0.0150	0.3409	0.04
1	2	0	1	8	8	0	1	0	-3.7	0.0247	0.3409	0.07
1	3	0	0	6	0	0	0	0	-1.2	0.3012	0.3409	0.88
2	1	1	0	12	8	1	0	0	-4.6	0.0101	0.1823	0.06
2	2	0	1	6	8	0	1	0	-3.3	0.0369	0.1823	0.20
2	3	0	0	10	0	0	0	0	-2	0.1353	0.1823	0.74
3	1	1	0	6	8	0	0	0	-3.9	0.0202	0.1495	0.14
3	2	0	1	12	4	0	1	0	-2.9	0.0550	0.1495	0.37
3	3	0	0	8	0	0	0	1	-2.6	0.0743	0.1495	0.50
4	1	1	0	6	4	0	0	0	-2.3	0.1003	0.1869	0.54
4	2	0	1	10	4	0	0	0	-3.3	0.0369	0.1869	0.20
4	3	0	0	10	0	0	0	1	-3	0.0498	0.1869	0.27
5	1	1	0	10	4	0	0	0	-3.1	0.0450	0.1334	0.34
5	2	0	1	8	4	0	0	0	-2.9	0.0550	0.1334	0.41
5	3	0	0	12	0	0	0	1	-3.4	0.0334	0.1334	0.25
6	1	1	0	12	4	1	0	0	-3	0.0498	0.1774	0.28
6	2	0	1	10	4	0	0	0	-3.3	0.0369	0.1774	0.21
6	3	0	0	12	0	0	0	0	-2.4	0.0907	0.1774	0.51
7	1	1	0	8	8	0	0	0	-4.3	0.0136	0.3259	0.04
7	2	0	1	12	8	0	1	0	-4.5	0.0111	0.3259	0.03
7	3	0	0	6	0	0	0	0	-1.2	0.3012	0.3259	0.92
8	1	1	0	8	4	1	0	0	-2.2	0.1108	0.2016	0.55
8	2	0	1	6	8	0	0	0	-4.1	0.0166	0.2016	0.08
8	3	0	0	8	0	0	0	1	-2.6	0.0743	0.2016	0.37

Intermediate computations …

t	j	asc1	asc2	A	B	C	D	E	asc1	asc2	A	B	C	D	E
1	1	0.04	0.00	0.44	0.35	0.04	0.00	0.00	0.04	0.07	6.32	0.93	0.04	0.07	0.00
1	2	0.00	0.07	0.58	0.58	0.00	0.07	0.00	0.04	0.07	6.32	0.93	0.04	0.07	0.00
1	3	0.00	0.00	5.30	0.00	0.00	0.00	0.00	0.04	0.07	6.32	0.93	0.04	0.07	0.00
2	1	0.06	0.00	0.66	0.44	0.06	0.00	0.00	0.06	0.20	9.30	2.06	0.06	0.20	0.00
2	2	0.00	0.20	1.21	1.62	0.00	0.20	0.00	0.06	0.20	9.30	2.06	0.06	0.20	0.00
2	3	0.00	0.00	7.42	0.00	0.00	0.00	0.00	0.06	0.20	9.30	2.06	0.06	0.20	0.00
3	1	0.14	0.00	0.81	1.08	0.00	0.00	0.00	0.14	0.37	9.20	2.55	0.00	0.37	0.50
3	2	0.00	0.37	4.42	1.47	0.00	0.37	0.00	0.14	0.37	9.20	2.55	0.00	0.37	0.50
3	3	0.00	0.00	3.97	0.00	0.00	0.00	0.50	0.14	0.37	9.20	2.55	0.00	0.37	0.50
4	1	0.54	0.00	3.22	2.15	0.00	0.00	0.00	0.54	0.20	7.85	2.93	0.00	0.00	0.27
4	2	0.00	0.20	1.97	0.79	0.00	0.00	0.00	0.54	0.20	7.85	2.93	0.00	0.00	0.27
4	3	0.00	0.00	2.66	0.00	0.00	0.00	0.27	0.54	0.20	7.85	2.93	0.00	0.00	0.27
5	1	0.34	0.00	3.38	1.35	0.00	0.00	0.00	0.34	0.41	9.68	3.00	0.00	0.00	0.25
5	2	0.00	0.41	3.30	1.65	0.00	0.00	0.00	0.34	0.41	9.68	3.00	0.00	0.00	0.25
5	3	0.00	0.00	3.00	0.00	0.00	0.00	0.25	0.34	0.41	9.68	3.00	0.00	0.00	0.25
6	1	0.28	0.00	3.37	1.12	0.28	0.00	0.00	0.28	0.21	11.58	1.95	0.28	0.00	0.00
6	2	0.00	0.21	2.08	0.83	0.00	0.00	0.00	0.28	0.21	11.58	1.95	0.28	0.00	0.00
6	3	0.00	0.00	6.14	0.00	0.00	0.00	0.00	0.28	0.21	11.58	1.95	0.28	0.00	0.00
7	1	0.04	0.00	0.33	0.33	0.00	0.00	0.00	0.04	0.03	6.29	0.61	0.00	0.03	0.00
7	2	0.00	0.03	0.41	0.27	0.00	0.03	0.00	0.04	0.03	6.29	0.61	0.00	0.03	0.00
7	3	0.00	0.00	5.55	0.00	0.00	0.00	0.00	0.04	0.03	6.29	0.61	0.00	0.03	0.00
8	1	0.55	0.00	4.40	2.20	0.55	0.00	0.00	0.55	0.08	7.84	2.86	0.55	0.00	0.37
8	2	0.00	0.08	0.49	0.66	0.00	0.00	0.00	0.55	0.08	7.84	2.86	0.55	0.00	0.37
8	3	0.00	0.00	2.95	0.00	0.00	0.00	0.37	0.55	0.08	7.84	2.86	0.55	0.00	0.37

Z matrix

t	j	asc1	asc2	A	B	C	D	E
1	1	0.20	-0.02	0.77	1.48	0.20	-0.02	0.00
1	2	-0.01	0.25	0.45	1.90	-0.01	0.25	0.00
1	3	-0.04	-0.07	-0.30	-0.88	-0.04	-0.07	0.00
2	1	0.22	-0.05	0.63	1.39	0.22	-0.05	0.00
2	2	-0.02	0.36	-1.48	2.67	-0.02	0.36	0.00
2	3	-0.05	-0.17	0.60	-1.78	-0.05	-0.17	0.00
3	1	0.32	-0.14	-1.18	2.00	0.00	-0.14	-0.18
3	2	-0.08	0.38	1.70	0.88	0.00	0.38	-0.30
3	3	-0.10	-0.26	-0.85	-1.80	0.00	-0.26	0.35
4	1	0.34	-0.14	-1.36	0.78	0.00	0.00	-0.20
4	2	-0.24	0.36	0.95	0.47	0.00	0.00	-0.12
4	3	-0.28	-0.10	1.11	-1.51	0.00	0.00	0.38
5	1	0.38	-0.24	0.19	0.58	0.00	0.00	-0.15
5	2	-0.22	0.38	-1.08	0.64	0.00	0.00	-0.16
5	3	-0.17	-0.21	1.16	-1.50	0.00	0.00	0.38
6	1	0.38	-0.11	0.22	1.08	0.38	0.00	0.00
6	2	-0.13	0.36	-0.72	0.93	-0.13	0.00	0.00
6	3	-0.20	-0.15	0.30	-1.40	-0.20	0.00	0.00
7	1	0.20	-0.01	0.35	1.51	0.00	-0.01	0.00
7	2	-0.01	0.18	1.05	1.37	0.00	0.18	0.00
7	3	-0.04	-0.03	-0.28	-0.58	0.00	-0.03	0.00
8	1	0.33	-0.06	0.12	0.85	0.33	0.00	-0.27
8	2	-0.16	0.26	-0.53	1.47	-0.16	0.00	-0.11
8	3	-0.33	-0.05	0.10	-1.73	-0.33	0.00	0.38

Fisher information matrix

	b1	b2	b3	b4	b5	b6	b7
b1	1.17	-0.41	-0.73	3.80	0.54	-0.07	-0.50
b2	-0.41	1.13	-0.07	3.97	-0.12	0.49	-0.37
b3	-0.73	-0.07	17.64	-5.41	0.52	0.68	0.62
b4	3.80	3.97	-5.41	47.35	1.84	2.50	-3.85
b5	0.54	-0.12	0.52	1.84	0.54	-0.01	-0.20
b6	-0.07	0.49	0.68	2.50	-0.01	0.49	-0.18
b7	-0.50	-0.37	0.62	-3.85	-0.20	-0.18	0.87

Expected covariance matrix

	b1	b2	b3	b4	b5	b6	b7
b1	16.05	13.48	0.25	-1.59	-5.79	-1.60	6.01
b2	13.48	13.37	0.19	-1.45	-3.92	-2.65	5.36
b3	0.25	0.19	0.09	0.01	-0.30	-0.30	0.06
b4	-1.59	-1.45	0.01	0.22	0.34	-0.05	-0.48
b5	-5.79	-3.92	-0.30	0.34	5.29	1.36	-1.75
b6	-1.60	-2.65	-0.30	-0.05	1.36	4.91	-0.72
b7	6.01	5.36	0.06	-0.48	-1.75	-0.72	4.15

Sample size requirements

Key concepts & study plan

Experimental design

Data collection & processing

Model specification & estimation

Interpretation & application

Sample size requirements

Minimum sample size requirement for each parameter

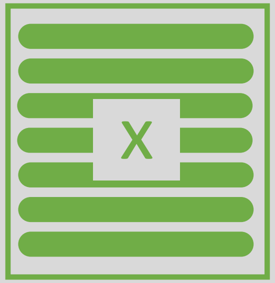
- Expected covariance matrix can be used to predict the minimum required sample size for statistically significant parameters

Sample size estimates

Model formulation

$U(A) = \beta x_A$
$U(B) = \beta x_B$
$U(C) = \beta x_C$

Experimental design



Parameter priors

β

Expected covariance matrix

$var(\beta | x)$

Sample size requirements

Minimum sample size requirement for each parameter

- Assuming a 5% significance level, desired t-ratio for parameter β_k is

$$|t_k| = \frac{|\beta_k|}{\text{se}(\beta_k \mid \mathbf{x}) / \sqrt{N_k}} \geq 1.96$$

- Re-arranging terms, corresponding sample size estimate is

$$N_k \geq \left(\frac{1.96 \cdot \text{se}(\beta_k \mid \mathbf{x})}{\beta_k} \right)^2$$

Sample size requirements

Example

- What minimum sample size is needed for each parameter to be statistically significant?
- 

$$\text{var}(\boldsymbol{\beta} \mid \mathbf{x}) = \begin{pmatrix} 0.36 & -0.14 & -0.07 \\ -0.14 & 0.79 & 0.06 \\ -0.07 & 0.06 & 0.12 \end{pmatrix}$$

$$\beta_1 = 0.15$$
$$\beta_2 = -0.13$$
$$\beta_3 = 0.32$$

$$N_1 \geq \left(\frac{1.96 \cdot \sqrt{0.36}}{0.15} \right)^2$$

$$N_2 \geq \left(\frac{1.96 \cdot \sqrt{0.79}}{-0.13} \right)^2$$

$$N_3 \geq \left(\frac{1.96 \cdot \sqrt{0.12}}{0.32} \right)^2$$

Sample size requirements

Example

- What minimum sample size is needed for each parameter to be statistically significant?
- 

$$\text{var}(\boldsymbol{\beta} \mid \mathbf{x}) = \begin{pmatrix} 0.36 & -0.14 & -0.07 \\ -0.14 & 0.79 & 0.06 \\ -0.07 & 0.06 & 0.12 \end{pmatrix}$$

$$\beta_1 = 0.15$$
$$\beta_2 = -0.13$$
$$\beta_3 = 0.32$$

$$N_1 \geq 61.5$$

$$N_2 \geq 179.6$$

$$N_3 \geq 4.5$$

Sample size requirements

Caveats of sample size estimates

- They are only meaningful with informative (non-zero) priors
- They are not exact (because priors are merely best guesses)
- They assume that a decision-maker faces all choice tasks in the design
  - Multiply with number of blocks in design to obtain actual sample size estimate
- They indicate minimum requirements
  - For more reliable parameter estimates (with higher *t*-ratios) you need a larger sample size

Sample size requirements

More information

- Rose, J.M. & Bliemer, M.C.J. (2013) Sample size requirements for stated choice experiments. *Transportation*, 40, 1021-1041.
- De Bekker-Grob, E.W., Donkers, B., Jonker, M.F. & Stolk, E.A. (2015) Sample size requirements for discrete choice experiments in healthcare: a practical guide. *The Patient*, 8, 373-384.