

# Designing Efficient but Unbalanced Choice Experiments

**John Henstridge**  
*Data Analysis Australia*  
**December 2010**

# Overview

---

## → Problem

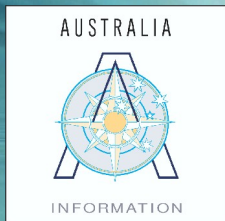
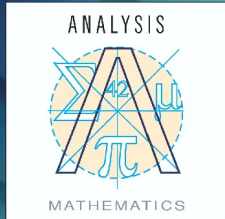
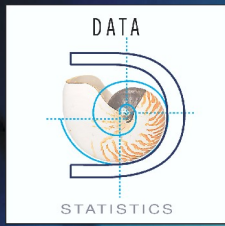
- Designing a choice questionnaire for stated travel preferences
- To be implemented across several modes
  - And to be unbalanced

## → Solution

- D-optimal design
- Random search procedures

## → Acknowledgments

- Many DAA staff but particularly Donna Hill



# Context

## → Planning of new train route

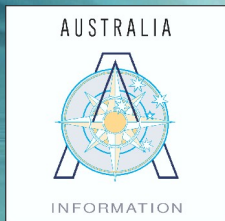
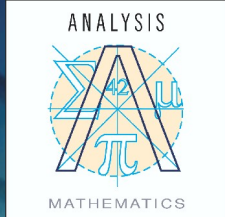
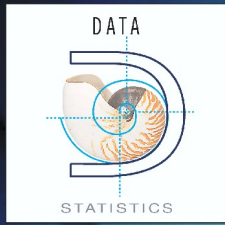
- ~200 kilometres
- More frequent, quicker
- Aim to take pressure of road

## → Preference modelling

- Route, frequency, speed, price, train *versus* car

## → Survey required

- Multiple modes
  - Mail, intercept, face-to-face



# Typical Choices

---

## → Route

→ Three train routes, plus bus

## → Frequency

→ Every 30 minutes, every hour etc

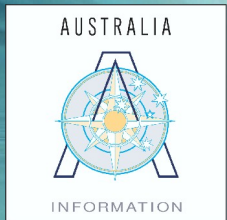
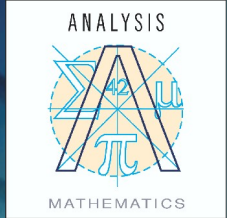
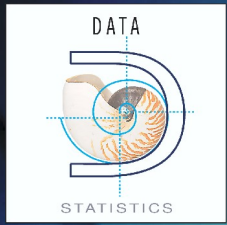
## → Speed - time of travel

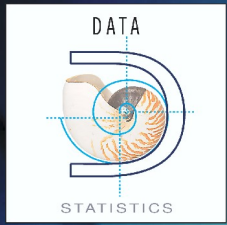
## → Price (fare)

## → Mode

→ Train, bus, car

- Car currently dominates!





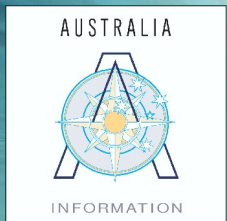
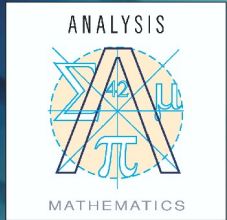
# Choice Question

→ **Choice set has several scenarios**

→ Question is to rank them or choose "most preferred"

→ Simplest form is to compare two

→ **Makes respondent think**



The following options include travel by train, coach and car. When considering the option of travelling by car, approximate costs and travel times have been provided based on travel between city centres. Coach travel times are also based on travel between city centres.

	Option A	Option B	Option C
<b>Mode/Route</b>	Red line	Car	Coach/Bus
<b>Cost</b>	\$29.70 per person (\$14.80 concession)	\$20.00 per car	\$26.00 per person (\$13.00 concession)
<b>Travel Time</b>	1½ hrs (90 mins)	2 hrs 20 mins	3 hrs 20 mins
<b>Frequency of Departure</b>	two hourly	N/A	twice a day

Q . Which of these three options would you be most likely to use? Option \_\_\_\_\_

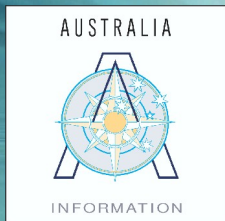
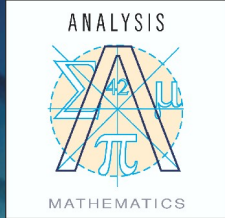
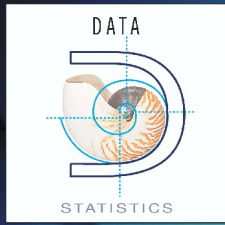
# Design

→ **Questionnaire typically has several such questions**

- Typically different choice sets presented to respondents
- Fractional design

→ **Ideally a balanced design?**

- E.g. Street, D. & Burgess, L.B. 2007, *The Construction of Optimal Stated Choice Experiments: Theory and Methods*



# Analysis

## → Multinomial response

→ Nested logit models

- Louviere, Henscher and Swait, 2000, *Stated Choice Methods*

## → Understanding binary choices

→  $P_{12} = \mathbf{Prob}$  choosing 1 over 2

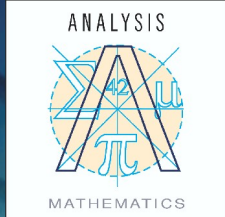
→  $\text{Logit}(P_{12}) = U_1 - U_2$

→ Linear model for  $U_1, U_2$

→  $U_1 - U_2 = (X_1 - X_2)\beta = \mathbf{X}\beta$

## → GLM like models

→ Degrade balance of design if effects significant



# Constraints

## → Some choices are obvious

→ If one option is faster, cheaper, more frequent (and otherwise the same) you would always choose it.

## → Some too different

→ Hard to compare

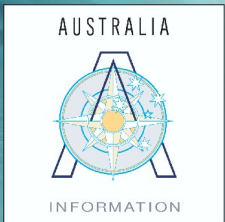
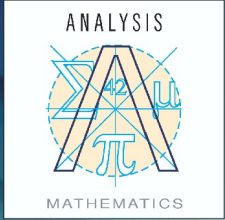
## → Choice set must be reasonable

→ At least one option someone might choose

- E.g. car might have to always be option

## → Different survey modes have different constraints

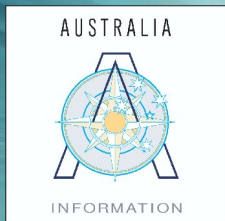
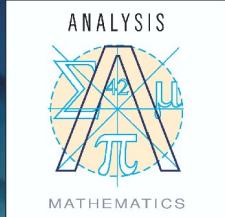
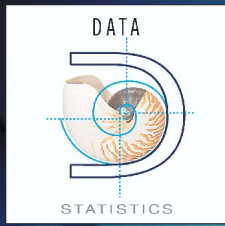
## → Lead to lack of balance

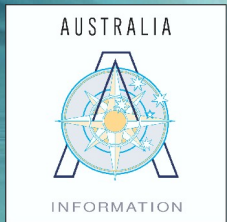
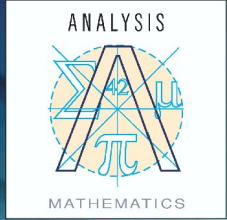
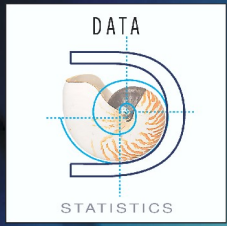


# Design Approach

---

- A. Generate design matrix  $X$  with all possible rows**
- B. Eliminate rows that conflict with constraints**
- C. Stratify if appropriate**
- D. Select subsets of required size within each stratum**
- E. Optimise the selection**





# What form of optimisation?

→ **Many possible rows to choose from**

→ Typically thousands

→ Enormous number of subsets

- E.g. 
$$\binom{1248}{32} = 8 \times 10^{98}$$

→ **What criterion?**

→ D-optimal principle

- Minimises product of parameter variances

→ **Combinatorial problem**

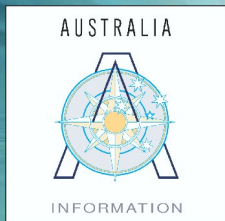
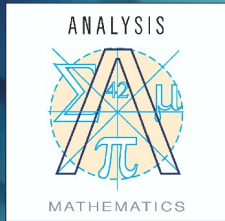
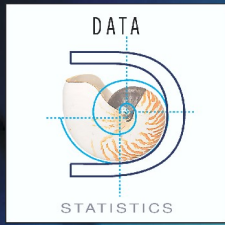
→ Difficult

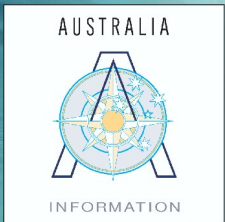
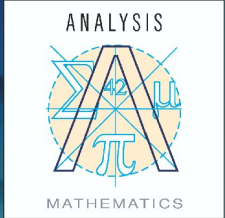
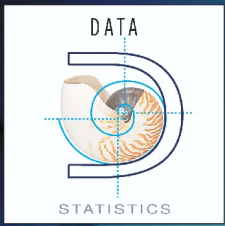
# Methods (1)

## → “Best random choice”

- Easy to implement
- No problems with local optima
- Terribly slow

```
best<-0
for(n in 1:BIGNUMBER) {
  i<-sample(SET,k)
  x<-X[i,]
  d<-det(t(x)%*%x)
  if(d>best){
    keep<-i
    best<-d
  }
}
```





# Methods (2)

## → Local search

→ Will find best *local* optimum

→ Requires concept of local – metric or topology

- Two samples of same size  $k$  are neighbours if they have  $k-1$  elements in common

→ Faster

```
keep<-sample(SET, k)
x<-X[keep, ]
best<-det(t(x)%*%x)
for(n in 1:BIGNUMBER) {
  i<-union(sample(keep, k-1),
            sample(setdiff(SET, keep), 1))
  x<-X[i, ]
  d<-det(t(x)%*%x)
  if(d>best){
    keep<-i
    best<-d
  }
}
```

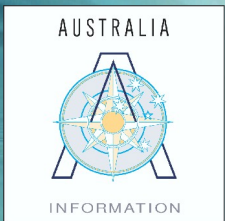
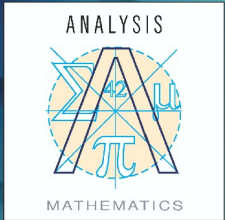
# Methods (3)

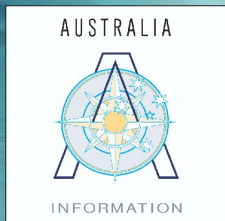
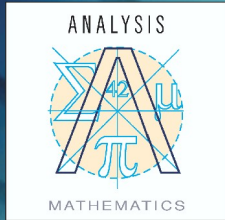
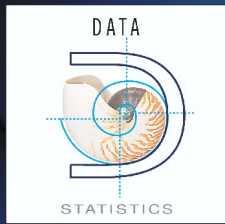
## → Narrowing search

→ Start with larger jumps – say 10 steps

→ Much faster, less likely to be trapped

```
keep<-sample(SET, k)
x<-X[keep,]
best<-det(t(x)%*%x)
for (r in 10:1) {
  for(n in 1:NOTSOBIGNUMBER) {
    i<-union(sample(keep, k-r),
              sample(setdiff(SET, keep), r))
    x<-X[i,]
    d<-det(t(x)%*%x)
    if(d>best){
      keep<-i
      best<-d
    }
  }
}
```





# Methods (4)

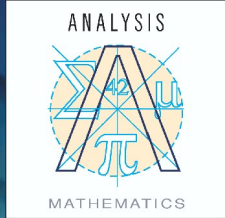
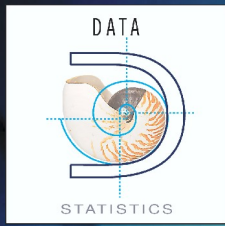
## → Simulated annealing

→ Local search with probabilistic acceptance

```
keep<-sample(SET,k)
x<-X[keep,]
best<-det(t(x)%*%x)
for(n in 1:BIGNUMBER) {
  i<-union(sample(keep,k-1),
            sample(setdiff(SET,keep),1))
  x<-X[i,]
  d<-det(t(x)%*%x)
  if(d>best-something(n)){
    keep<-i
    best<-d
  }
}
```

→ something requires tuning

- Critically, it contains a random component



# Methods (5)

---

## → Multiple narrowing search

- Number of good independent start points
  - Followed by narrowing search
- Similar properties of annealing methods
  - Similar speed but less tuning required

## → Other methods

- Tabu search
  - Eccleston, J. and Chan, B. (1998). Design algorithms for correlated data, *COMPSTAT 1998*

# Outcomes

## → Methods work

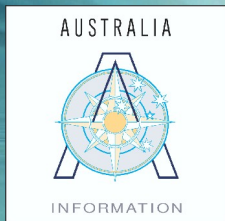
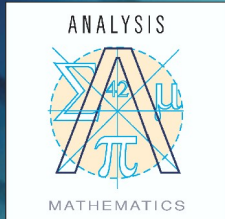
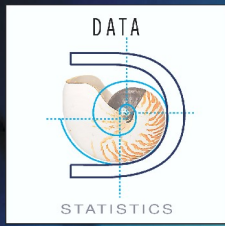
→ Flexible enough to adapt to different contexts

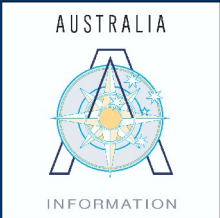
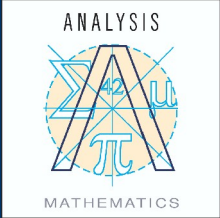
- Easy to program
- Computationally feasible

→ Typically several hours

→ Provide designs that seem reasonable

- Worked in practice
- Optimisation of design reduced sample sizes





# Thank you



[daa.com.au](http://daa.com.au)

STRATEGIC  
INFORMATION  
CONSULTANTS