Week 10: *DOE & Power Analysis*

EMSE 6035: Marketing Analytics for Design Decisions

John Paul Helveston

November 01, 2023

Before we start, re-install {cbcTools}

```r
remotes::install_github("jhelvy/cbcTools")
```

Week 10: *DOE & Power Analysis*

1. Design of Experiment

2. Design Efficiency

3. Power Analysis

Week 10: *DOE & Power Analysis*

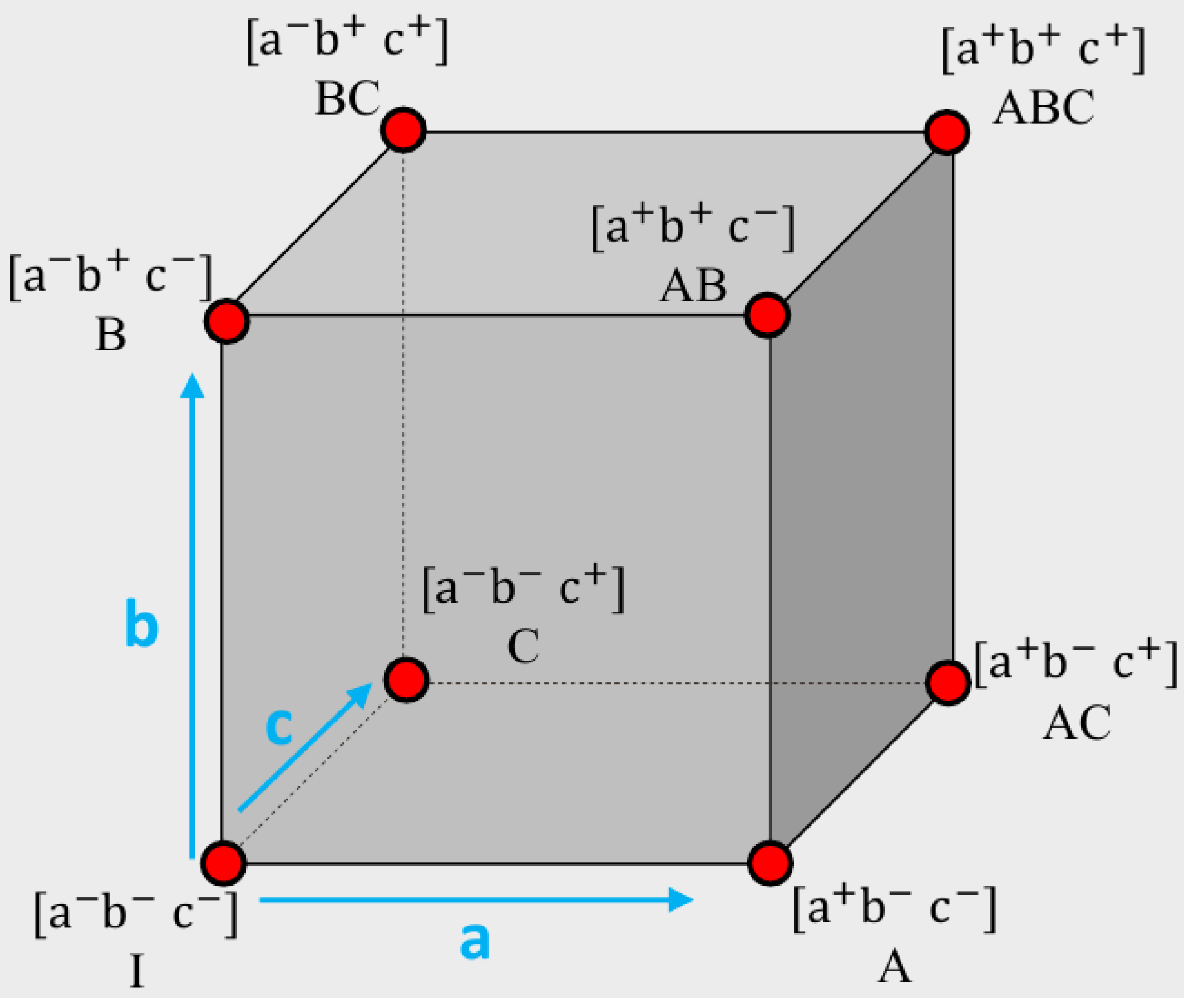
1. Design of Experiment

2. Design Efficiency

3. Power Analysis

Main & Interaction Effects

Full design space for 3 effects: A, B, C



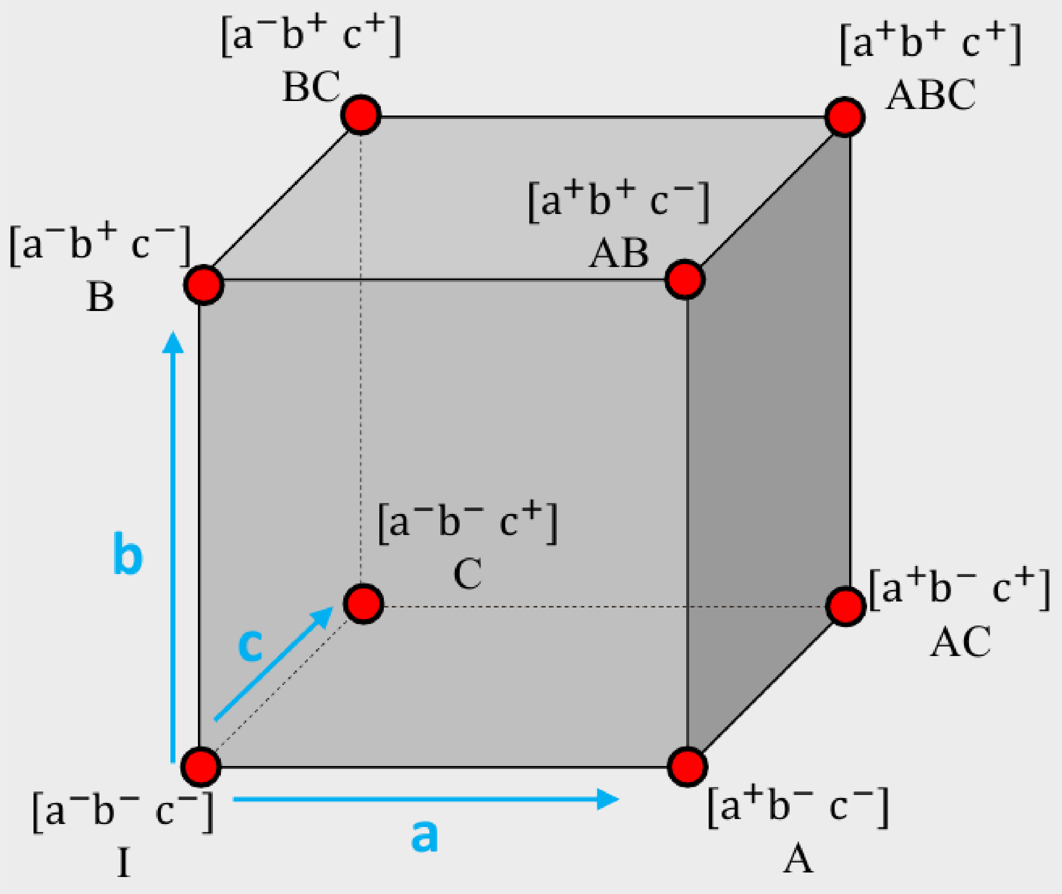
Full design space for 3 effects: A, B, C

Example: *Cars*

A: Electric? (Yes+ or No-)

B: Warranty? (Yes+ or No-)

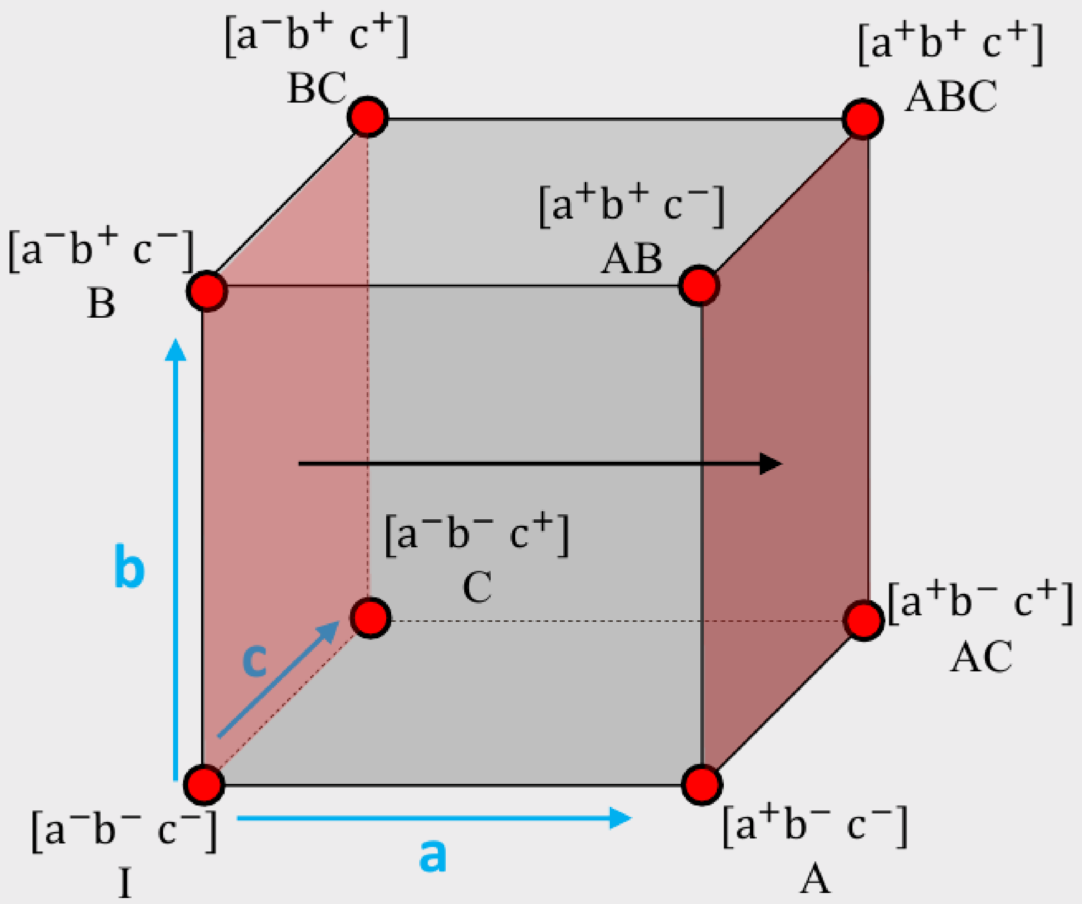
C: Ford? (Yes+ or No-)



Main Effects

$$ME(a) = \left(\frac{A + AB + AC + ABC}{4}\right) - \left(\frac{I + B + C + BC}{4}\right)$$

(A: Electric? Yes+ or No-)

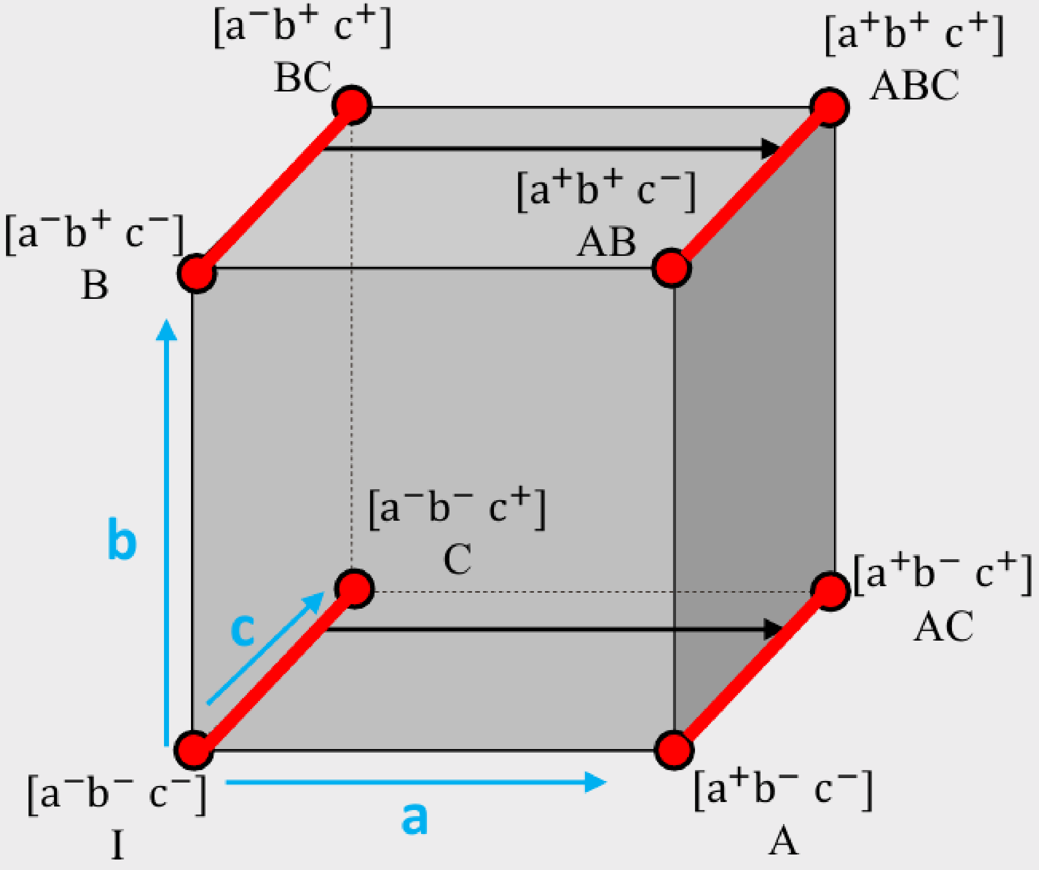


Interaction Effects

$$INT(ab) =$$

$$\frac{1}{2}\left[\left(\frac{AB + ABC}{2}\right) - \left(\frac{B + BC}{2}\right)\right] -$$

$$\frac{1}{2}\left[\left(\frac{A + AC}{2}\right) - \left(\frac{I + C}{2}\right)\right]$$



Example: Wine Pairings

meat	wine
fish	white
fish	red
steak	white
steak	red

Main Effects

1. `meat`: **Fish** or **Steak**?
2. `wine`: **Red** or **White**?

Example: Wine Pairings

meat	wine
fish	white
fish	red
steak	white
steak	red

Main Effects

1. `meat`: **Fish** or **Steak**?
2. `wine`: **Red** or **White**?

Interaction Effects

1. `meat*wine`: **Red** or **White** wine *with **Steak***?
2. `meat*wine`: **Red** or **White** wine *with **Fish***?

Open `interactions.qmd`

Fractional vs Full Factorial Designs

Full Factorial Design

Example: *Cars*

A: Electric? (Yes+ or No-)

B: Warranty? (Yes+ or No-)

C: Ford? (Yes+ or No-)

```r
library(cbcTools)

profiles <- cbc_profiles(
    electric = c(1, 0),
    warranty = c(1, 0),
    ford     = c(1, 0)
)

profiles
```

```
#>   profileID electric warranty ford
#> 1         1        1        1    1
#> 2         2        0        1    1
#> 3         3        1        0    1
#> 4         4        0        0    1
#> 5         5        1        1    0
#> 6         6        0        1    0
#> 7         7        1        0    0
#> 8         8        0        0    0
```

Full Factorial Design

Balanced?

All levels appear an equal number of times.

Orthogonal?

All pairs of levels appear together an equal number of times.

```r
library(cbcTools)

profiles <- cbc_profiles(
    electric = c(1, 0),
    warranty = c(1, 0),
    ford     = c(1, 0)
)

profiles
```

```
#>   profileID electric warranty ford
#> 1         1        1        1    1
#> 2         2        0        1    1
#> 3         3        1        0    1
#> 4         4        0        0    1
#> 5         5        1        1    0
#> 6         6        0        1    0
#> 7         7        1        0    0
#> 8         8        0        0    0
```

Fractional Factorial Design

Balanced?

All levels appear an equal number of times.

Orthogonal?

All pairs of levels appear together an equal number of times.

```r
profiles[c(1, 3, 5, 6),]
```

```
#>   profileID electric warranty ford
#> 1         1        1        1    1
#> 3         3        1        0    1
#> 5         5        1        1    0
#> 6         6        0        1    0
```

Comparing Full and Fractional Factorial Designs

Open `balance-orthogonality.qmd`

Practice Question 1

Consider the following experiment design

a	b	c	Effect
+	-	-	A
-	+	-	B
+	-	+	AC
-	+	+	BC

a) Is the design balanced? Is is orthogonal?

b) Write out the equation to compute the main effect for a, b, and c.

c) Are any main effects confounded? If so, what are they confounded with?

Week 10: *DOE & Power Analysis*

1. Design of Experiment
2. Design Efficiency
3. Power Analysis

A simple conjoint experiment about *cars*

Attribute	Levels
Brand	GM, BMW, Ferrari
Price	$20k, $40k, $100k

Design: 9 choice sets, 3 alternatives each

```
Attribute counts:

brand:
  GM    BMW   Ferrari
  10     11     6

price:

 20k  40k 100k
  9    9   9
```

```
Pairwise attribute counts:

brand & price:

          20k 40k 100k
  GM        3   0    7
  BMW       4   5    2
  Ferrari   2   4    0
```

A simple conjoint experiment about *cars*

Attribute	Levels
Brand	GM, BMW, Ferrari
Price	$20k, $40k, $100k

Design: 90 choice sets, 3 alternatives each

```
Attribute counts:

brand:
  GM     BMW    Ferrari
  92     80      98

price:

  20k  40k 100k
  91   84   95
```

```
Pairwise attribute counts:

brand & price:

          20k 40k 100k
  GM      31  31  30
  BMW     25  25  30
  Ferrari 35  28  35
```

Bayesian D-efficient designs

Maximize information on "Main Effects" according to priors

Attribute	Levels	Prior
Brand	GM, BMW, Ferrari	0, 1, 2
Price	$20k, $40k, $100k	0, -1, -4

$$v_j = 1\delta^{\text{BMW}} + 2\delta^{\text{Ferrari}} - 1\delta^{\text{40k}} - 4\delta^{\text{100k}}$$

Bayesian D-efficient designs

Maximize information on "Main Effects" according to priors

Attribute	Levels	Prior
Brand	GM, BMW, Ferrari	0, 1, 2
Price	$20k, $40k, $100k	0, -1, -4

```
Attribute counts:

brand:
  GM     BMW    Ferrari
  93     90      86

price:

  20k  40k 100k
  97   93   78
```

```
Pairwise attribute counts:

brand & price:

          20k 40k 100k
  GM      52  41  0
  BMW     30  30  30
  Ferrari 15  22  49
```

Negative of the hessian evaluated at a set of parameters is called the **"Information Matrix"**

$$\boldsymbol{I}(\boldsymbol{\beta}) = -\nabla^2_{\boldsymbol{\beta}} \ln(\mathcal{L})$$

"D-optimal" designs attempt to minimize the "D-error" of a design

$$D = \left| \boldsymbol{I}(\boldsymbol{\beta}) \right|^{-1/p}$$

where p is the number of coefficients in the model

Finding Efficient Designs

Open `design-efficiency.qmd`

Your Turn

20:00

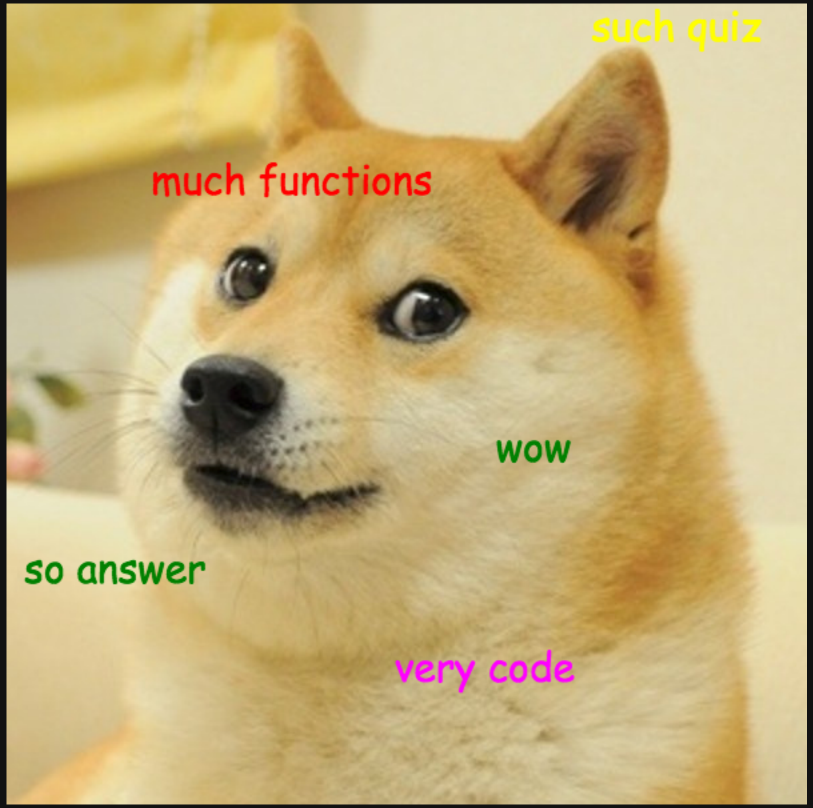
1. Individually, create a Bayesian D-efficient fractional factorial survey design. Inspect the attribute balance and overlap.

2. Compare your results with your teammates.

Quiz 4

Link is in the #class channel

10:00



Week 10: *DOE & Power Analysis*

1. Design of Experiment

2. Design Efficiency

3. Power Analysis

How many respondents do I need?

How many respondents do I need

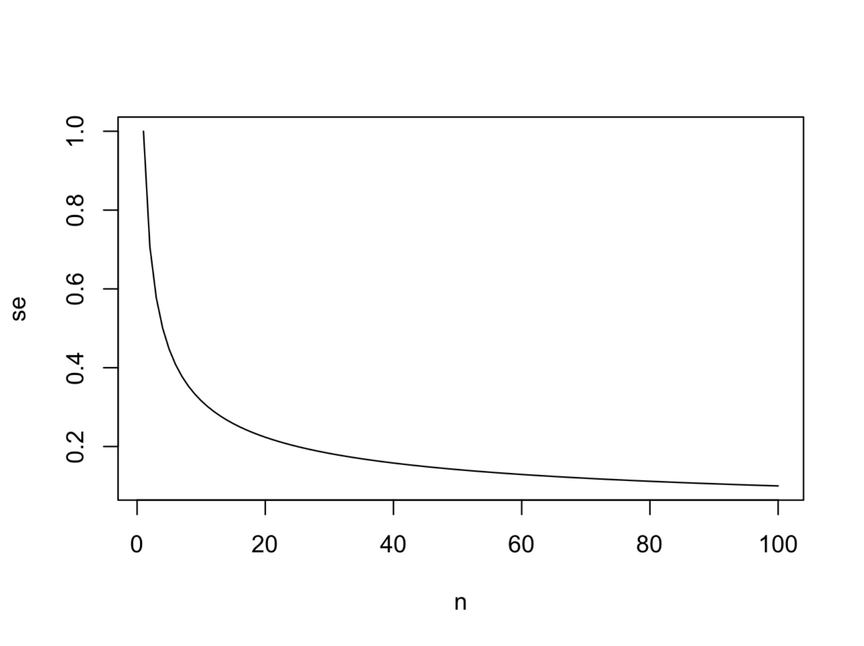
to get X level of precision on β?

Standard errors are inversely related to $\sqrt{N}$

```r
n <- seq(100)
se <- 1/sqrt(n)
plot(n, se, type = "l")
```

Standard errors also decrease with:

- Fewer attributes
- Fewer levels in each categorical attribute
- More questions per respondent



Using {cbcTools}, we can run simulations to determine the necessary sample size for a specific model

Open `powerAnalysis.qmd`

Your Turn

20:00

Individually:

1. Using the survey design you created in the last practice, conduct a power analysis to determine the necessary sample size to achieve a 0.05 significance level on your parameter estimates.

2. Compare your results with your teammates.