Inside experimental designs

Basics and ground rules for DCM and conjoint

Steven Struhl

About experimental designs

- Choice models, conjoint and MaxDiff are based on **experimental designs** [1]
  - Designs ensure that we can measure all effects accurately when we vary many attributes together
  - They also greatly reduce the number of items (marketplaces or conjoint cards)from the number of possible combinations of all factors
    - For instance, suppose you had a product with:
      - 6 attributes, each having 3 levels, and one attribute with 6 levels
    - This would mean that you could have 3 x 3 x 3 x 3 x 3 x 3 x 6 or some **4374** possible variations on this product
      - With experimental design, we can accurately get the value for all 4374 possible variations using **only 18 product descriptions**
    - Suppose you have a product with 18 two-level attributes
      - This would give you 2^18 or **262,144** possible combinations
      - You can measure all these possible **using only 20 product descriptions**

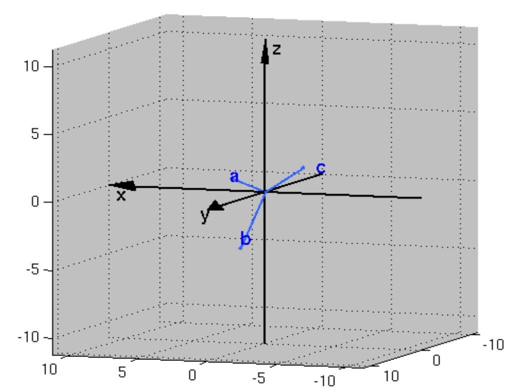
Experimental designs took off on the farm, trying to get better crops. It's a long story. Maybe another time.

[1]This mainly covers formal designed experiments. We get to the type called random designs later.

The designs we use: Orthogonal and D-optimal

- Many types of experimental designs have been devised—just a few are highly important for choice models, conjoint and MaxDiff
  - The type traditionally used in choice models and conjoint is called an **orthogonal fractional factorial** design
    - The design discussed on the next page is "fractional factorial."
    - To review: designs like this let us estimate many the worth or utility of many attribute levels, using very few marketplaces or product descriptions
    - However, they sacrifice something to get this great efficiency
      - They are not designed to measure *interactions* among attributes
      - Interactions will be coming up shortly
  - Another design, aided by computers, is very similar to the fractional factorial type—the **D-optimal design**
    - These may provide real benefits in reducing the size of a design vs. a standard fractional factorial

Not to confuse things unduly, but axes at right angles (like x, y and z here) also are called orthogonal—and they indeed have zero correlation



Inside fractional factorial experimental designs

- Each ***attribute*** appears as a ***variable*** (column) in the design (as would be the case with any design)
- The levels are encoded (in this case starting with zero; some programs start with one)
  - e.g., three levels for one attribute would appear in the design as 0, 1, 2 in one column
- Each card or marketplace or product profile will be one row of the design
  - Reading across the row will give the levels of each attribute, in the order specified

	Attribute A	Attribute B	Attribute C
Card 1	0	0	0
Card 2	0	0	1
Card 3	0	1	0
Card 4	0	1	1
Card 5	1	0	0
Card 6	1	0	1
Card 7	1	1	0
Card 8	1	1	1

- Each attribute level appears the same number of times with each other attribute level
  - That is, all **pairwise combinations** are covered, but the design may not cover all possible three-way combinations of attribute levels)

Converge Analytic

More about designs: orthogonal means no correlations

- Note that the way in which each attribute is varied (from one card, or marketplace, to the next) has absolutely no correlation with the way in which any of the other attributes vary

- **No correlations** means that the way in which any given attribute varies cannot have an influence on the value we get for another attribute

  - This is all that **orthogonal** means: no correlations between the variations in any of the attributes

Correlations

		A	B	C
A	Pearson Correlation	1.000	.000	.000
	Sig. (2-tailed)	.	1.000	1.000
	N	8	8	8
B	Pearson Correlation	.000	1.000	.000
	Sig. (2-tailed)	1.000	.	1.000
	N	8	8	8
C	Pearson Correlation	.000	.000	1.000
	Sig. (2-tailed)	1.000	1.000	.
	N	8	8	8

Sorry: none of this involved!

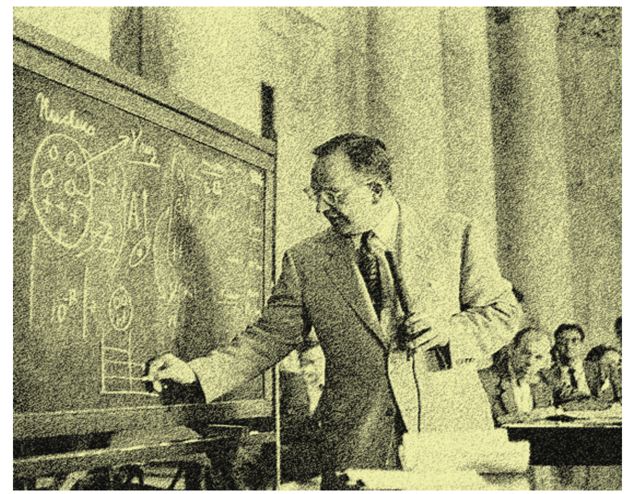
A more difficult fractional factorial design

- Here we have 5 attributes. One has 4 levels (0 to 3), one has 3 levels, and 2 have 2 levels
  - This requires 16 cards or marketplaces
- This also is a much more difficult balancing act (and not recommended for your spare time)

	Attribute A	Attribute B	Attribute C	Attribute D	Attribute E
Card 1	3	2	1	1	0
Card 2	2	2	0	0	0
Card 3	1	1	0	0	1
Card 4	1	2	1	0	1
Card 5	3	0	0	0	1
Card 6	0	2	0	1	1
Card 7	0	0	1	0	0
Card 8	2	0	0	1	1
Card 9	2	0	1	1	1
Card 10	2	1	1	0	0
Card 11	3	0	1	0	1
Card12	1	0	1	1	0
Card 13	3	1	0	1	0
Card 14	1	0	0	1	0
Card 15	0	0	0	0	0
Card 16	0	1	1	1	1

Once again, all the correlations among the different attributes are zero

		A	B	C	D	E
A	Pearson Correlation	1.000	.000	.000	.000	.000
	Sig. (2-tailed)	.	1.000	1.000	1.000	1.000
	N	16	16	16	16	16
B	Pearson Correlation	.000	1.000	.000	.000	.000
	Sig. (2-tailed)	1.000	.	1.000	1.000	1.000
	N	16	16	16	16	16
C	Pearson Correlation	.000	.000	1.000	.000	.000
	Sig. (2-tailed)	1.000	1.000	.	1.000	1.000
	N	16	16	16	16	16
D	Pearson Correlation	.000	.000	.000	1.000	.000
	Sig. (2-tailed)	1.000	1.000	1.000	.	1.000
	N	16	16	16	16	16
E	Pearson Correlation	.000	.000	.000	.000	1.000
	Sig. (2-tailed)	1.000	1.000	1.000	1.000	.
	N	16	16	16	16	16



Back when you could have a fun time talking about zero correlations. (This is not your author.)

Design standards: How many scenarios or products

- We need more cards or marketplaces to measure more attributes and levels
- Here's a quick general check for minimum design size (design saturation, or when the design gets completely full)
  - (Number of attributes X number of levels) - number of attributes + 1
  - **Example:** 8 attributes, three with four levels, 5 with 2 levels:
    - (3 x 4) + (5 x 2) = 22
    - subtract 8 = 14
    - add 1 = 15
  - Therefore, select the smallest design that requires at least 15 cards (this will in fact be a 16 scenario or card design)
  - **Note**
  - Some say you need to add 3 instead of 1,

    or multiply the subtracted total (13, in this case)

    by 1.1 and round up
    - However, adding one works fine

Poor designs can hurt you

Getting more from data—the world of HB analysis

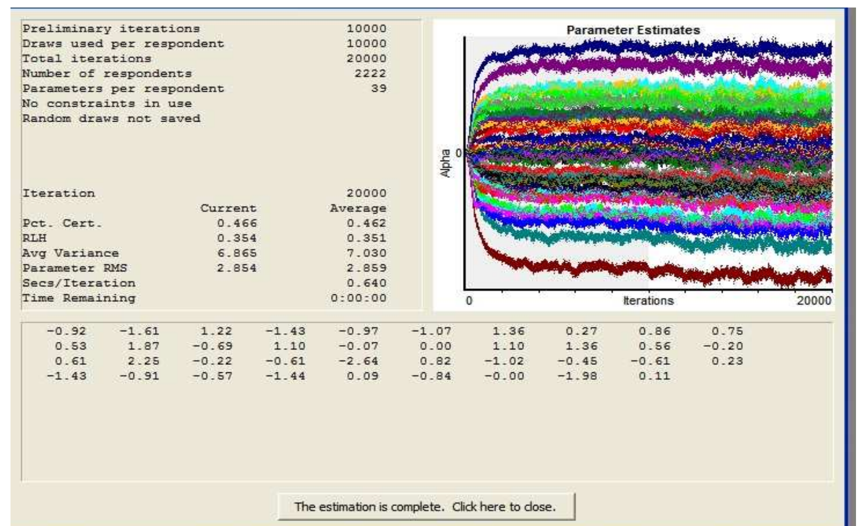
- Hierarchical Bayesian or HB analysis can stretch the limits of standard designs
  - You can get **3 or 4 times more information** reliably—amazing but true
  - However, we still need to know 3 or 4 times **more than what** amount
- Fixed design programs usually use catalogues of designs rather than designing "from scratch"
- No design may exist that approaches the theoretical lower limit in size
  - Therefore, if you used fixed designs, the "best" design you can find may greatly exceed the minimum for saturation (or the design being "full")
  - **D-optimal** designs may help out here
  - However, even D-optimal designs cannot exceed **saturation**
- With standard conjoint, if you end up needing more than 18 cards, the design is too large
  - You must trim back!
- This is not so with DCM
  - As mentioned, thanks to HB, we now can measure as much as a respondent can stand in a study

All the latest HB hits playing here

Converge Analytic

What is HB analysis?

- This method fills in data that is scant or missing for a respondent by repeatedly borrowing estimates from other respondents
  - That is, it keeps sampling other respondents and storing the values that those with information have
  - It usually runs 20,000 or more times for each respondent, keeping a running average of its estimates
    - It may or may not compare the respondent to the sample it is drawing and make adjustments based on their similarity
  - Estimates will settle down to steady values (or **converge**) if you have set up the problem correctly
    - If you have not done so, maybe not
    - A solution that does not converge usually means errors in setup, data collection or coding of values
  - It gives your PC more of a workout than almost anything else
  - You will wait for a complicated DCM run to finish
- Amazingly, **all this borrowing works—** and we get very accurate estimates

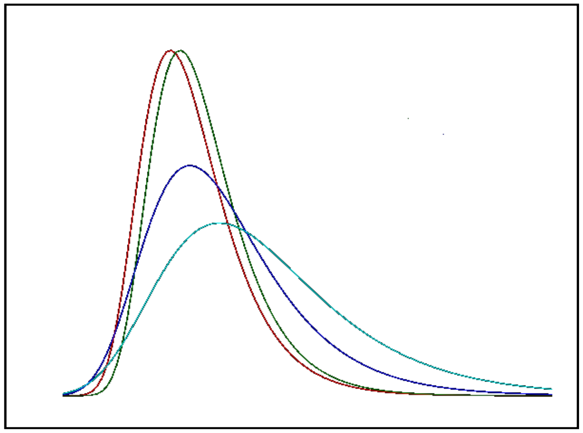


How values vary from 0 to 20,000 estimates (and looking mostly stable at the end)

From design size to sample size

- You need enough sample for the analysis to provide reliable results
- Errors for DCM tend to be tighter than the errors around sample percentages—meaning we may get away with a slightly smaller sample for a given level of precision
- For **standard analysis** (no HB analysis added), getting to a minimum good sample is fairly simple
  - Start by thinking of **125 (full/complete) respondents** as the minimum
    - Both for conjoint analysis and for DCM where people choose one item
    - This means 125 respondents per group you want to analyze separately
  - Next, look at the number of attributes and levels in your design
    - Use your favorite formula to determine how many scenarios/cards that will require
  - Then work up the sample as follows . . .

This is the distribution of errors we expect for choice models. This is narrower than our old friend the "normal" distribution. Narrower error means more precise measurements.



An example of getting to sample size

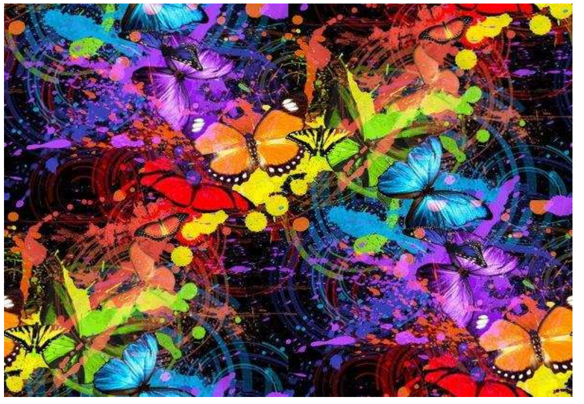
- Say you have 12 attributes with 3 levels
- You need 12 * 3 (or 36) - 13 (+1) or 24 cards or scenarios
- This gives you one **replication** or the equivalent of one full respondent.
  - Now, suppose you worry that your respondents will tired easily and so you want to give each person in the survey 8 scenarios to rate
  - That works out to exactly one-third of a "full respondent" apiece
    - This means you will **need to interview 375** people to get the equivalent of **125** full respondents
- If respondents allocate (e.g., over the next 10 patients) and they are a fairly homogeneous population, you can get by with 75 per group
- Remember, **with HB analysis** and a good size experiment, you can **get 3 to 4** times more out of each respondent
  - Or, you most likely could get by only 125—getting practically all the power you would with the 375 sample
  - That is a vast improvement!

On the Web illustrating "vast improvement." One never knows.

Converge Analytic

A note on “random designs”

- A software company (Sawtooth software) has proposed “random designs”[1] for discrete choice (which they call choice-based-conjoint)
- A computer mixes up the attributes and levels and gives a random combination to each respondent
- With enough respondents, this should cover every combination of attributes and levels
  - Some really good analysts say this works well
- There could be some concerns about how well it pans out with HB (Hierarchical Bayesian) analysis
  - That method fills in spotty data by “borrowing” repeatedly from samples of other respondents
  - It is not clear what it is borrowing if everybody is doing something different from everybody else
- You rely on everything coming out in the wash
  - In fact, it may well do so

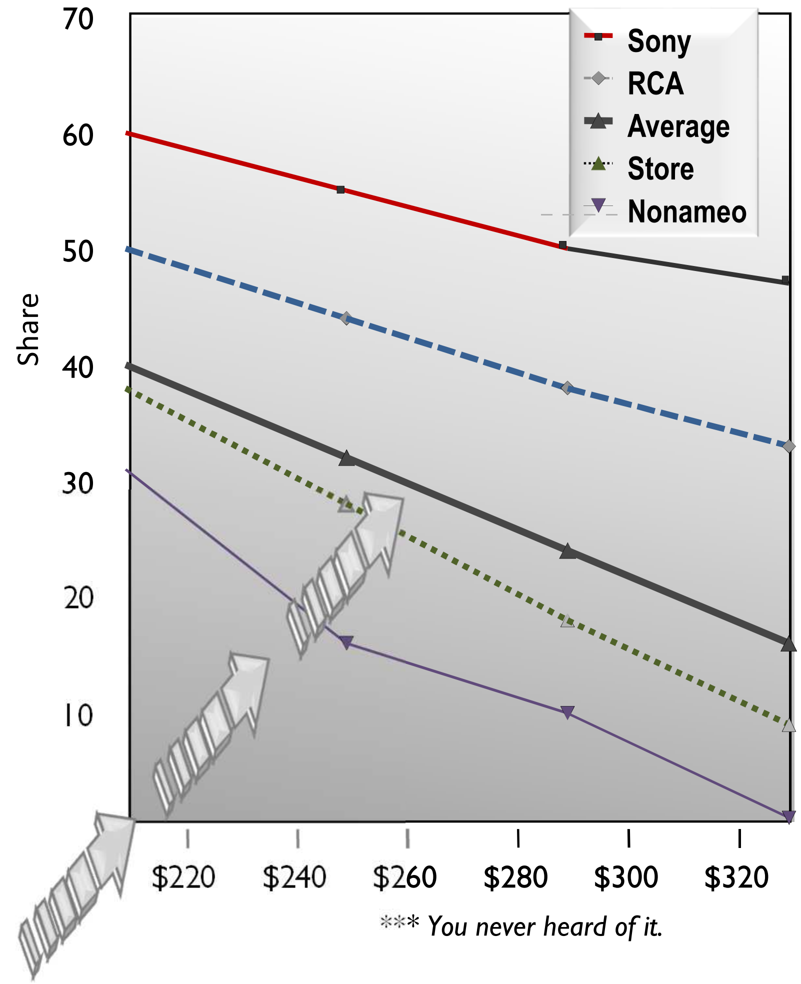


Another design billed as random

[1] *This does seem like an impossible combination of words*

Concluding side-bar: What is an interaction?

- **Interaction** has practical and statistical meanings
  - Practically, this means:
    - To determine a variable we care about (like market share) we must understand how two (or more) other variables influence each other
    - That is, we must know how both variables behave together to get accurate readings of share changes.
      - Knowing just one is not enough
- **Example.** Suppose we have 4 brands of televisions: Sony, RCA, Store Brand, and Nonameo***
  - Suppose each of these brands could be sold at any of 4 prices:
    - $209, $249, $289, and $329
  - If (e.g.) Sony sells better at all prices than the other brands, then brand and price **interact**
  - With conjoint style designs, you must specify that brand and price interact to see the different price vs. share response patterns for the 4 brands
    - Otherwise, you get an average ("generic") price curve that does not fit most of the brands well
- **DCM eliminates the need for this brand-price interaction** with attributes specific to each brand



*** *You never heard of it.*

Overall average

Converge Analytic

Why we care about interactions

- Interactions can blow out your design or blow up your measurements
- When added to a model, they multiply the number of terms
  - So for instance, 3 brand and 4 prices, handled with an interaction term, adds **12 more terms** to your model
    - You actually need a total of 19 terms—3 for brand, 4 for price and 12 for the interaction
  - You would do better giving each brand had its own price attributes
    - That would be 12 attribute levels also (4 prices x 3 brands)
    - But that would be all!
    - We do not need to measure brand—it comes along “for free” as a constant in the choice model
    - So we save 7 terms in the model—and get accurate, direct measurement in the bargain

We have to be careful about multiplying

Converge Analytic

Questions? Comments?

There is more--We are just touching the edges

Dr. Steven Struhl

smstruhl@convergeanalytic.com
smstruhl@gmail.com

847-624-2268

Converge Analytic