



Inflection Point Analytics Inc.

Designed Experiments for Web & Email Marketing

Alison Burnham
November 25, 2008

Outline

1. Background
2. Experimental Design Theory
3. Experimental Design Implementation Notes
4. Email Example
5. Web application
6. Extensions
7. References



Background

Early Designed Experiments

- In 1747, while serving as surgeon on HM Bark *Salisbury*, James Lind, the ship's surgeon, carried out a controlled experiment to develop a cure for scurvy.
- Lind selected 12 men from the ship, all suffering from scurvy, and divided them into six pairs, giving each group different additions to their basic diet for a period of two weeks.
- The treatments were all remedies that had been proposed at one time or another. They were:
 - A quart of cider every day
 - Twenty five gutts of *elixir vitriol* three times a day upon an empty stomach,
 - One half-pint of seawater every day
 - A mixture of garlic, mustard, and horseradish in a lump the size of a nutmeg
 - Two spoonfuls of vinegar three times a day
 - Two oranges and one lemon every day.



Background

Early Designed Experiments



- The men who had been given citrus fruits recovered dramatically within a week. One of them returned to duty after 6 days and the other became nurse to the rest.
- The others experienced some improvement, but nothing was comparable to the citrus fruits, which were proved to be substantially superior to the other treatments.
- In this study his subjects' cases "were as similar as I could have them", that is he provided strict entry requirements to reduce extraneous variation.
- The men were paired, which provided replication.
- From a modern perspective, the main thing that is missing is randomized allocation of subjects to treatments.



Background



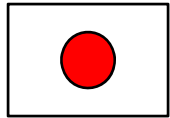
R. A. Fisher -

- Wrote initial defining book "Design of Experiments" 1935

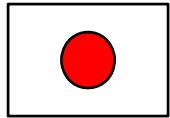


W. E. Deming -

- best known for his work in Japan from 1950 onward



- taught top management how to improve design, product quality, testing and sales through various methods, including the application of statistical experimentation methods.



G. Taguchi

- Japanese engineer and statistician
- Worked with Deming, extended experimental concepts



USA - Quality Revolution 1980's

- "Designed Experiments" major tool in "Quality Toolbox"



Background

Marketing

- A/B split experiments are well entrenched in marketing - people are careful when testing alternatives to make allocations random, have adequate sample sizes etc
- Issues
 - Testing many things at one time becomes challenging - number of groups required is incremental
 - Testing sequentially becomes time consuming - pace of marketing decision making too fast to wait
- Moving to more complex designs has started in marketing but still low adoption / awareness
- Web & Email are starting to produce large enough sample sizes to make large scale experimentation possible



Experimental Design Theory

- Want to test impact of new creative for an email on the clickthrough to the company website
- Do an A/B split - form two segments
- Send out first segment in the morning
- Send out second segment in the afternoon
- CTR is much higher for the second segment
- But is it the creative or time of day?
- These two factors are confounded



Experimental Design Theory

- To complicate further:
 - A number of things were varied in the creative between option 1 and 2
 - Background colour - light vs dark
 - Font on offer
 - Wording around offer
 - Overall length of the email
 - Arguments now happening around which creative element caused the effect
 - They're all confounded!



Experimental Design Theory

- Idea is to test multiple factors at the same time without testing “all combinations”.
- Example - two factors -
 1. Introductory copy
Levels : included vs. not included
 2. Background highlight color
Levels included vs. not included



Experimental Design Theory

- All combinations would be 4 versions:
 1. introductory copy, background highlight color
 2. no introductory copy, background highlight color
 3. introductory copy, no background highlight color
 4. no introductory copy, no background highlight color



Experimental Design Theory

- For 3 factors = $2 \times 2 \times 2 = 8$ versions
- For 4 factors = $2 \times 2 \times 2 \times 2 = 16$ versions
- For 5 factors = $2 \times 2 \times 2 \times 2 \times 2 = 32$ versions
-
-
- for 13 factors = $2 \times 2 \times \dots \times 2 = 8192$ versions
- and so on...



Experimental Design Theory

- In our example - add another factor:
 - Short vs Long Template
- Production has said for the project budget it can only create 4 versions
- Do we drop a factor to “test later”?
- Or is there any way to test all 3 in 4 runs?



Experimental Design Theory

- All Combinations

Run	Introductory Copy	Background Highlight Color	Short vs Long Template
1	Yes	Yes	Short
2	Yes	Yes	Long
3	Yes	No	Short
4	Yes	No	Long
5	No	Yes	Short
6	No	Yes	Long
7	No	No	Short
8	No	No	Long

- We need to pick 4 of these for our test....



Experimental Design Theory

- The right 4 runs:

Run	Introductory Copy	Background Highlight Color	Short vs Long Template
1	Yes	Yes	Short
4	Yes	No	Long
6	No	Yes	Long
7	No	No	Short

- For any given factor if you compare the two levels, the other two factors are balanced
- If we run these 4 versions then we can test each of the three factors “cleanly”
- Called a “fractional factorial experiment”



Experimental Design Theory

- The results:

Run	Introductory Copy	Background Highlight Color	Short vs. Long Template	Result (CTR)
1	Yes	No	Short	15.0 %
2	Yes	Yes	Long	17.2 %
3	No	No	Long	13.4 %
4	No	Yes	Short	17.6 %

- Introductory copy impact is:

Average at Yes - Average at No

$$\frac{(15.0 + 17.2)}{2} - \frac{(13.4 + 17.6)}{2} = 0.6 \% \text{ (Higher CTR for Introductory Copy)}$$



Experimental Design Theory

- The results:

Run	Introductory Copy	Background Highlight Color	Short vs. Long Template	Result (CTR)
1	Yes	No	Short	15.0 %
2	Yes	Yes	Long	17.2 %
3	No	No	Long	13.4 %
4	No	Yes	Short	17.6 %

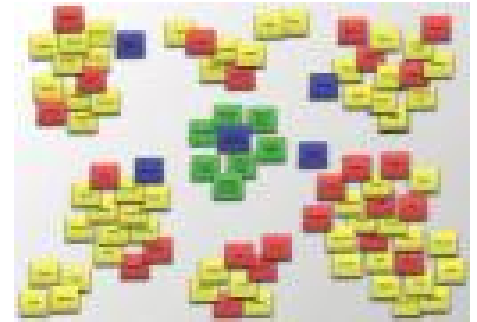
- Background Highlight Color impact is

Average at Yes - Average at No

$$\frac{(17.2 + 17.6)}{2} - \frac{(15.0 + 13.4)}{2} = 3.1 \% \text{ (Higher CTR for Bkgd Highlight Color)}$$



Experimental Design Implementation Notes



- Factors:
 - Factors are obtained through group brainstorming including as diverse a team as you can assemble
 - Get as many ideas as you can!
 - Sometimes it's the unexpected factor that can be a breakthrough
 - Run the experiment under as many noise conditions as you can

Definition "Noise" - something you can't control going forward e.g. weather, competitive campaigns



Experimental Design Implementation Notes

- Response:
 - While you're at it - consider collecting multiple responses
 - Make sure response is clearly defined and agreed to by the group
- Results
 - Remember - sometimes no impact is the best result
 - imagine finding out that paper stock has no impact on DM response!?
 - It may save you money to show something does not affect the results
 - Things can be statistically significant but not practically significant - don't get carried away by the math!



Email Example

- A B2B newsletter is being analyzed and is found to have very low click through rates relative to client expectations.
- Quantitative research sessions had obtained subscriber feedback on template layout and design.
- Creative team members also had ideas on how the template could be changed to encourage subscriber interaction.
- Following some team brainstorming it was determined that testing should be undergone in order to optimize the template layout and design.



Response

- The following response variables were measured for the experiment:
 - Distinct click through rate as a percentage of open
 - the number of customers who clicked at least once on the email as a percentage of the numbers of customers who have opened the email
 - Total click through rate as a percentage of open*
 - the count of total clicks on the email divided by the numbers of customers who have opened the email



Factors

- The following factors were considered for this experiment
 - Overall look & feel
 - Use of images
 - Image size
 - Number of articles
 - Placement of articles
 - Layout of page
 - Color fields
 - Subject line content
 - Article order
 - Navigation bar menu options
 - Navigation bar order
 - Background color
 - Use of captions / no captions
 - Font size
 - Overall amount of copy
 - Email Length



Factors

- The following factors were selected to be tested within this experiment:
 - ✓ Subscriber Engagement Segment
 - ✓ Template Length
 - ✓ Introduction Copy
 - ✓ Template Background Color
 - ✓ Promo Box Header Highlight



Experimental Design

- The following 8 runs were proposed:

Version #	Segment	Template	Promo Box Highlight	Intro	Background Color	Sample
1	Low/Med	Short	Yes	Yes	No	25,000
2	Low/Med	Short	Yes	No	Yes	25,000
3	Low/Med	Long	No	Yes	Yes	25,000
4	Low/Med	Long	No	No	No	25,000
5	High	Short	No	No	No	10,000
6	High	Short	No	Yes	Yes	10,000
7	High	Long	Yes	No	Yes	10,000
8	High	Long	Yes	Yes	No	10,000



Experimental Plan

- The subscriber base was split into 8 groups to be treated consistently throughout 3 deployments of the newsletter
- Preliminary analysis after one month,
- Final analysis after all 3 months complete



Results

- Results were very consistent across all 3 months of the experiment
- Shorter was better - shorter template showed distinctly higher click through (both total & distinct)
- Highlighting the promo boxes does not seem to enhance engagement - the templates with no highlighting performed consistently better
- Intro Copy and Background Color did not appear to impact engagement



Results : Detail

- Distinct Clicks

One factor showed as significant for Distinct Clicks (%O):

1. Template Length - shorter length template saw higher number of customers who clicked at least once

- Total Clicks

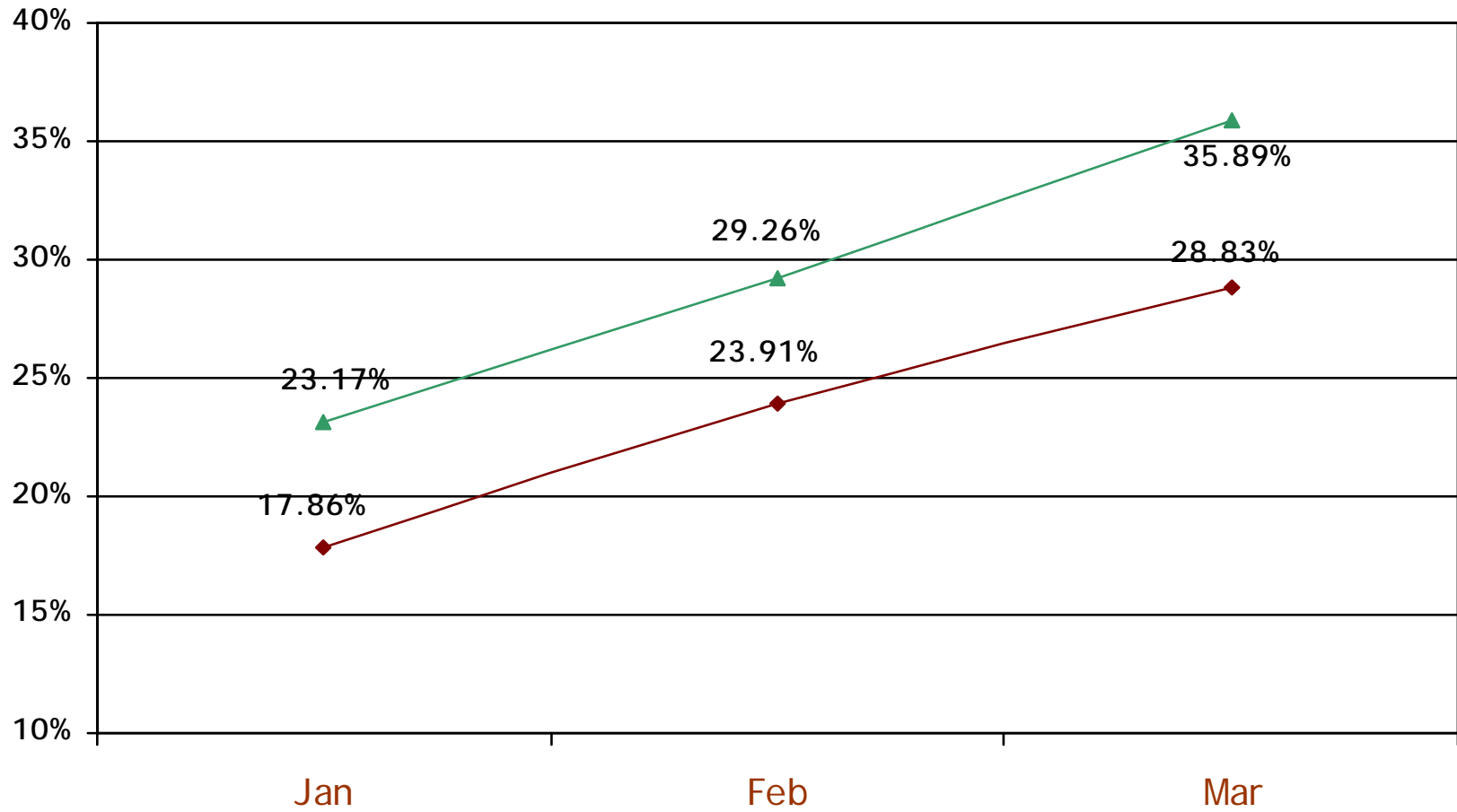
Two factors showed as significant for Total Clicks (%O):

1. Template Length - shorter length template saw higher total clicks
2. Right Promo Box Highlight - no highlighting saw higher total clicks



Results

Total Clicks : Promo Box Highlight



Web Application

- Theory exactly the same, issue is mechanism for showing different versions of the website to different audiences (i.e. how to enable randomization)
- Can do via login if site is restricted - truly randomized by subscriber
- If site is not restricted can do work around by rotating versions - need to ensure that time of day and day of week are considered as factors as well so nothing is confounded with timing



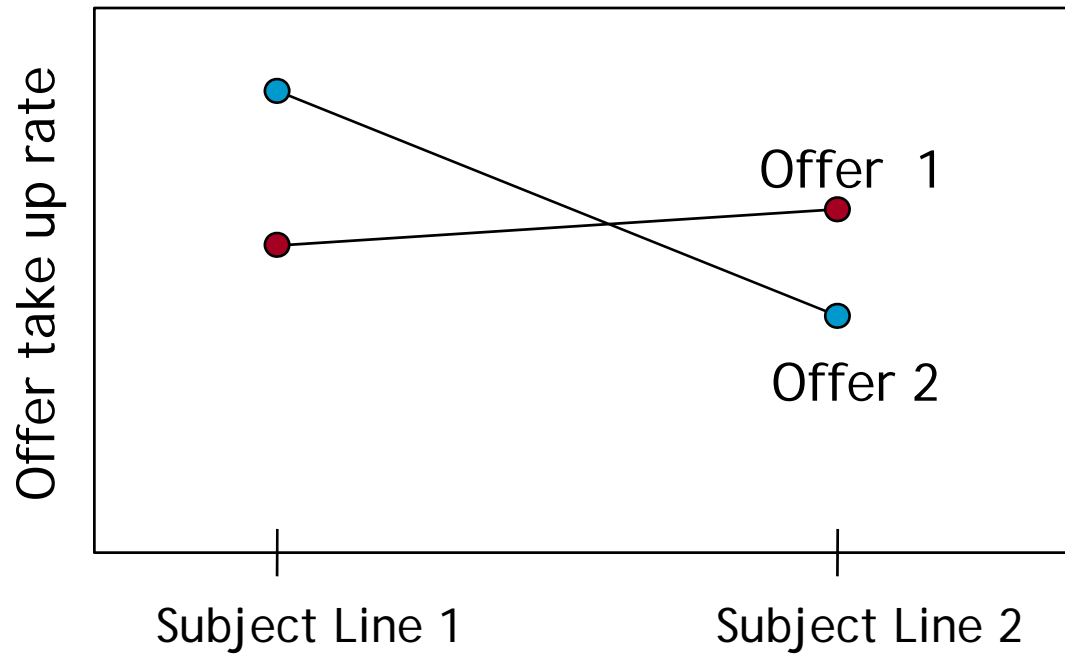
Extensions

- More than 5 factors
 - The experiment we showed had 5 factors in 8 runs - could have done as many as 7.
 - For 2 level factors - can do up to $n-1$ factors in a 2^n experiment (e.g. 7 in 8 runs, 15 in 16 runs).
- Other numbers of levels
 - There are designs that combine 2 level, 3 level, 5 level etc factors
 - Can find experimental design books like recipe books.



Extensions

- Interactions



- Which offer should you use ?



Summary

- Lot of potential within marketing for “smarter experimentation”
- Focus in tight times on “test & learn”, “pilot projects”
- When they were new in manufacturing - case studies shown where savings in the millions were obtained
- Can we translate that into marketing - what is a 1% lift in response worth?
- Math behind designs often reasonably simple and/or available in commercial statistics packages
- Or you could talk to your friendly neighborhood statistician!



References

1. Statistics for Experimenters by George Box, Stuart Hunter and William G. Hunter.
2. Testing 1-2-3 Experimental Design and Applications in Marketing and Service Organizations, Johannes Ledolter & Arthur J. Swersey
3. Business & Industrial Statistical Research Group (BISRG) University of Waterloo
4. www.marketingexperiments.com
5. <http://www.statease.com/articles.html>

