LATIN SQUARE DESIGNS:

BLOCKS ARE USEFUL… WHAT IF THERE ARE TWO BLOCK FACTORS?

-- HOW TO **REMOVE THE VARIATION FOR THESE TWO SOURCES FROM THE ERROR VARIATION**

EXAMPLE

A courier company is interested in deciding between five brands (D, P, F, C and R) of car for its next purchase of fleet cars.

- The brands are all comparable in purchase price.
- The company wants to carry out an experimental study that will enable them to compare the brands with respect to operating costs.
- For this purpose they select five drivers.
- In addition the study will be carried out over a five week period.

Latin Square Designs (Note: Sudoku is a special case!)

Selected Latin Squares

3 x 3

A	B	C
B	C	A
C	A	B

4 x 4

A	B	C	D
B	A	D	C
C	D	B	A
D	C	A	B

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

A	B	C	D
B	D	A	C
C	A	D	B
D	C	B	A

A	B	C	D
B	A	D	C
C	D	A	B
D	C	B	A

5 x 5

A	B	C	D	E
B	A	E	C	D
C	D	A	E	B
D	E	B	A	C
E	C	D	B	A

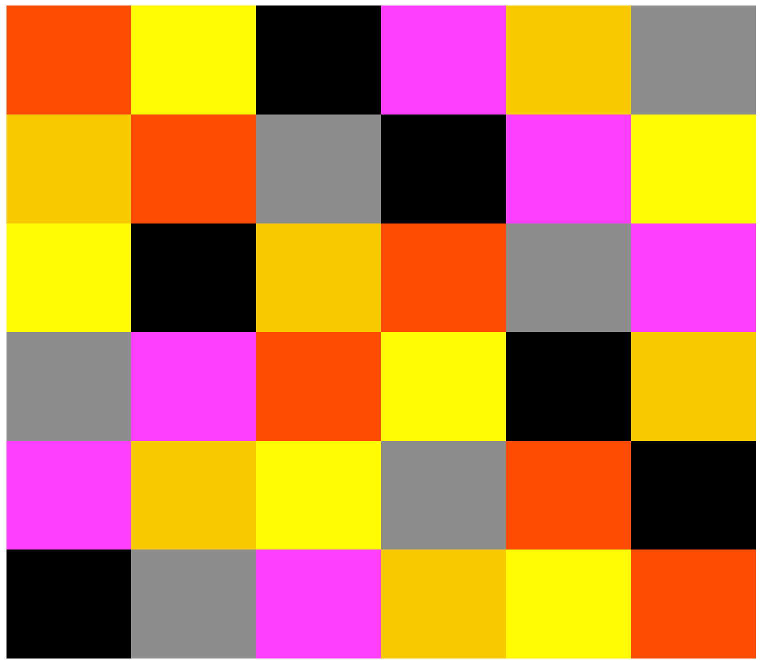
6 x 6

A	B	C	D	E	F
B	F	D	C	A	E
C	D	E	F	B	A
D	A	F	E	C	B
E	C	A	B	F	D
F	E	B	A	D	C

The Latin Square Design gets its name from the fact that we can write it as a square with Latin letters to correspond to the treatments.

NOTE:

Each object appears once and only once in each row and each column.
There are many such squares satisfying with a fixed size.



A Latin Square

DEFINITION

A Latin square is a square array of objects (letters A, B, C, …) such that each object appears once and only once in each row and each column. Example - 4 x 4 Latin Square (*in particular, a standardized Latin square*).

A	**B**	**C**	**D**
B	**C**	**D**	**A**
C	**D**	**A**	**B**
D	**A**	**B**	**C**

TWO BLOCKING VARIABLES: THE LATIN SQUARE DESIGN

Columns

Rows

1	2	3			t
2	3	1			
3	1	2			
⋮					
t					

Row, Column→ two ‘block’ factors

All t treatments (from one factor) appear once in each row and each column

In a Latin square You have three factors:

Treatments (*t*) (letters A, B, C, …)

Rows (*t*)

Columns (*t*)

- The number of treatments = the number of rows = the number of columns = t.
- The row-column treatments are represented by cells in a t x t matrix.
- The treatments are assigned to row-column combinations using a Latin-square arrangement

EXAMPLE

A courier company is interested in deciding between five brands (D, P, F, C and R) of car for its next purchase of fleet cars.

- The brands are all comparable in purchase price.
- The company wants to carry out a study that will enable them to compare the brands with respect to operating costs.
- For this purpose they select five drivers (=Rows).
- In addition the study will be carried out over a five week period (Columns = weeks).

- Each week a driver is assigned to a car using randomization and a Latin Square Design.
- The average cost per mile is recorded at the end of each week and is tabulated below:

Drivers		Week 1	Week 2	Week 3	Week 4	Week 5
	1	5.83 D	6.22 P	7.67 F	9.43 C	6.57 R
	2	4.80 P	7.56 D	10.34 C	5.82 R	9.86 F
Drivers	3	7.43 F	11.29 C	7.01 R	10.48 D	9.27 P
	4	6.60 R	9.54 F	11.11 D	10.84 P	15.05 C
	5	11.24 C	6.34 R	11.30 P	12.58 F	16.04 D

In practice, <u>Randomization</u> should be adhered:
Choose a standardized Latin square, and then
randomly permute the columns, randomly permute the rows.

The Model for a Latin Experiment

$$y_{ij(k)} = \mu + \tau_k + \rho_i + \gamma_j + \varepsilon_{ij(k)}$$

No interaction between rows, columns and treatments

$i = 1,2,\ldots, t \quad j = 1,2,\ldots, t \quad k = 1,2,\ldots, t$

$y_{ij(k)}$ = the observation in i^{th} row and the j^{th} column receiving the k^{th} treatment

μ = overall mean

τ_k = the effect of the k^{th} treatment

ρ_i = the effect of the i^{th} row

γ_j = the effect of the j^{th} column

$\varepsilon_{ij(k)}$ = random error

Variation due to drivers or weeks will be effectively removed from the error variation.

- A ***Latin Square*** experiment is assumed to be a **three**-factor experiment.
- The factors are ***rows***, ***columns*** and ***treatments***.
- It is assumed that there is ***no interaction*** between rows, columns and treatments.
- The degrees of freedom for the interactions is used to estimate error.

The ANOVA Table for a Latin Square Experiment

Source	S.S.	d.f.	M.S.	F	p-value
Treat	SS_{Tr}	t-1	MS_{Tr}	MS_{Tr}/MS_E	
Rows	SS_{Row}	t-1	MS_{Row}	MS_{Row}/MS_E	
Cols	SS_{Col}	t-1	MS_{Col}	MS_{Col}/MS_E	
Error	SS_E	(t-1)(t-2)	MS_E		
Total	SS_T	$t^2 - 1$			

The statistical analysis (ANOVA) is much like
the analysis for the RCBD (Randomized Complete Block Design).

The Anova Table for Example

Source	S.S.	d.f.	M.S.	F	p-value
Week	51.17887	4	12.79472	16.06	0.0001
Driver	69.44663	4	17.36166	21.79	0.0000
Car	70.90402	4	17.72601	22.24	0.0000
Error	9.56315	12	0.79693		
Total	201.09267	24			

GRAECO-LATIN SQUARE DESIGNS– TWO MAIN FACTORS, TWO BLOCK FACTORS OR, ONE MAIN FACTOR AND THREE BLOCK FACTORS

THE GRAECO-LATIN SQUARE DESIGN - AN EXAMPLE

A researcher is interested in determining the effect of two factors

- the percentage of *Lysine* in the diet and
- percentage of *Protein* in the diet

have on *Milk Production* in cows.

Previous similar experiments suggest that ***interaction*** between the two factors is negligible.

For this reason it is decided to use a Graeco-Latin square design to experimentally determine the two effects of the two factors (*Lysine* and *Protein*).

Seven levels of each factor is selected

- 0.0(A), 0.1(B), 0.2(C), 0.3(D), 0.4(E), 0.5(F), and 0.6(G)% for *Lysine* and
- 2(α), 4(β), 6(χ), 8(δ), 10(ε), 12(ϕ) and 14(γ)% for *Protein*).
- Seven animals (cows) are selected at random for the experiment which is to be carried out over seven three-month periods.

DEFINITION

- A **Graeco-Latin** square consists of two Latin squares (one using the letters A, B, C, … the other using Greek letters α, β, χ, …).
- A **Graeco-Latin square (Euler square, orthogonal Latin squares)** of order *n* over two sets *S* and *T*, each consisting of *n* symbols, is an $n \times n$ arrangement of cells, each cell containing an ordered pair (*s*,*t*), where *s* is in *S* and *t* is in *T*, such that every row and every column contains each element of *S* and each element of *T* exactly once, and that no two cells contain the same ordered pair.
- ***Example:*** a 7 x 7 Graeco-Latin Square

Aα	**Bε**	**Cβ**	**Dϕ**	**Eχ**	**Fγ**	**Gδ**
Bβ	**Cϕ**	**Dχ**	**Eγ**	**Fδ**	**Gα**	**Aε**
Cχ	**Dγ**	**Eδ**	**Fα**	**Gε**	**Aβ**	**Bϕ**
Dδ	**Eα**	**Fε**	**Gβ**	**Aϕ**	**Bχ**	**Cγ**
Eε	**Fβ**	**Gϕ**	**Aχ**	**Bγ**	**Cδ**	**Dα**
Fϕ	**Gχ**	**Aγ**	**Bδ**	**Cα**	**Dε**	**Eβ**
Gγ	**Aδ**	**Bα**	**Cε**	**Dβ**	**Eϕ**	**Fχ**

A Graeco-Latin Square is the used to assign the 7 X 7 combinations of levels of the two factors (*Lysine* and *Protein*) to a period and a cow. The data is tabulated on below:

		Period						
		1	2	3	4	5	6	7
Cows	1	304 (A^{α})	436 (B^{ε})	350 (C^{β})	504 (D^{ϕ})	417 (E^{χ})	519 (F^{γ})	432 (G^{δ})
	2	381 (B^{β})	505 (C^{ϕ})	425 (D^{χ})	564 (E^{γ})	494 (F^{δ})	350 (G^{α})	413 (A^{ε})
	3	432 (C^{χ})	566 (D^{γ})	479 (E^{δ})	357 (F^{α})	461 (G^{ε})	340 (A^{β})	502 (B^{ϕ})
	4	442 (D^{δ})	372 (E^{α})	536 (F^{ε})	366 (G^{β})	495 (A^{ϕ})	425 (B^{χ})	507 (C^{γ})
	5	496 (E^{ε})	449 (F^{β})	493 (G^{ϕ})	345 (A^{χ})	509 (B^{γ})	481 (C^{δ})	380 (D^{α})
	6	534 (F^{ϕ})	421 (G^{χ})	452 (A^{γ})	427 (B^{δ})	346 (C^{α})	478 (D^{ε})	397 (E^{β})
	7	543 (G^{γ})	386 (A^{δ})	435 (B^{α})	485 (C^{ε})	406 (D^{β})	554 (E^{ϕ})	410 (F^{χ})

The Model for a Graeco-Latin Experiment

$$y_{ij(kl)} = \mu + \tau_k + \lambda_l + \rho_i + \gamma_j + \varepsilon_{ij(kl)}$$

$i = 1,2,\ldots, t$ $\quad$ $j = 1,2,\ldots, t$

$k = 1,2,\ldots, t$ $\quad$ $l = 1,2,\ldots, t$

$y_{ij(kl)}$ = the observation in i^{th} row and the j^{th} column receiving the k^{th} Latin treatment and the l^{th} Greek treatment

μ = overall mean

τ_k = the effect of the k^{th} Latin treatment

λ_l = the effect of the l^{th} Greek treatment

ρ_i = the effect of the i^{th} row

γ_j = the effect of the j^{th} column

$\varepsilon_{ij(k)}$ = random error

No interaction between rows, columns, Latin treatments and Greek treatments

GENERALIZATIONS

Latin (Hyper)cube of order *n* :

For three (or more) sets *S* and *T, and R*, each consisting of *n* symbols, Latin cube is an $n \times n \times n$ arrangement of cells, each cell containing an ordered pair (*s*,*t*,*r*), where *s* is in *S* and *t* is in *T*, such that …

- Can handle more blocking variables

LATIN RECTANGLE: BY STACKING *M* LATIN SQUARES (HERE, M=2)

	Week 1	Week 2	Week 3	Week 4
Store 1	B	C	D	A
Store 2	A	B	C	D
Store 3	D	A	B	C
Store 4	C	D	A	B
Store 5	D	B	C	A
Store 6	C	A	B	D
Store 7	A	D	B	C
Store 8	B	D	A	C

We may want to include eight stores instead of four in our marketing experiment.

FACTORS AT THREE OR MORE LEVELS

THREE VS. TWO LEVELS?

- Two levels: the lowest level designs that can identify factor effects (low → high level what happens?)
- The main drawback of two level designs: they can only describe linear (straight line) trends in the response
- Three-level designs is the lowest level design that can capture curvature in the response function → This case, <u>quadratic</u> trends
- Price to pay? → More runs required

YOU SHOULD FIRST SEE..

- **… whether the order of levels matters or not (nominal vs. ordinal).**
- **E.g., ordinal factor**
  - Wealth level-low, medium, high
  - Education level, …
- **How about ethnicity (having three levels)?**

Cubic ‘trend’ does not carry any practical interpretation.

EXAMPLE: SALES OF APPLE JUICE

- **Question: Impact of price and display on sales?**
- **Price**
  - Low level: -1, the cost to the market
  - High level: 1, the retail price recommended by price manual
  - Reduced level: 0, halfway between the above two prices
- **Display**
  - Low level: -1, a reduced display space
  - High level: 1, extended display space
  - Mid level: 0, normal display space
- **Will conduct 3^2 full factorial experiment with one replication**

DATA (BASED ON DESIGN TABLE)

Display	Price	Sales
-1	-1	40.8
-1	-1	34.2
0	-1	44.2
0	-1	53.5
1	-1	91.5
1	-1	70.5
-1	0	32
-1	0	31.4
0	0	50.2
0	0	34.9
1	0	85.7
1	0	59.3
-1	1	9
-1	1	18
0	1	24.9
0	1	24.9
1	1	55.9
1	1	31.9

ANALYSIS

- **ANOVA (in this case Two-way, but Three-way or more factors possible) can find significant effects.**
- **In this example, there are three levels for each *continuous* factor, so one can examine whether curvatures are in present; that is, is there a quadratic trend in effect-response relationship?**
- **Such a question belongs to the realm of more general topic termed as *Response Surface Design*.**
  - Explore the functional relationship between the response and the factors, e.g., quadratic, cubic, etc.
  - Goal is to find optimal point(s) that minimize/ maximize the response variable.

ANALYSIS

- Proceed with full factorial ANOVA – one can see the main effects are significant.
- Then, carry out Response Surface to see if there exist quadratic trends – in this example, they are statistically insignificant.

* Similar steps can be taken for a factorial experiment wit multiple levels.

- For examples, two factors with two levels, and one factor with three levels: _____ factorial experiment
- ANOVA → provides significant effects

3-LEVEL *FRACTIONAL* FACTORIAL DESIGNS

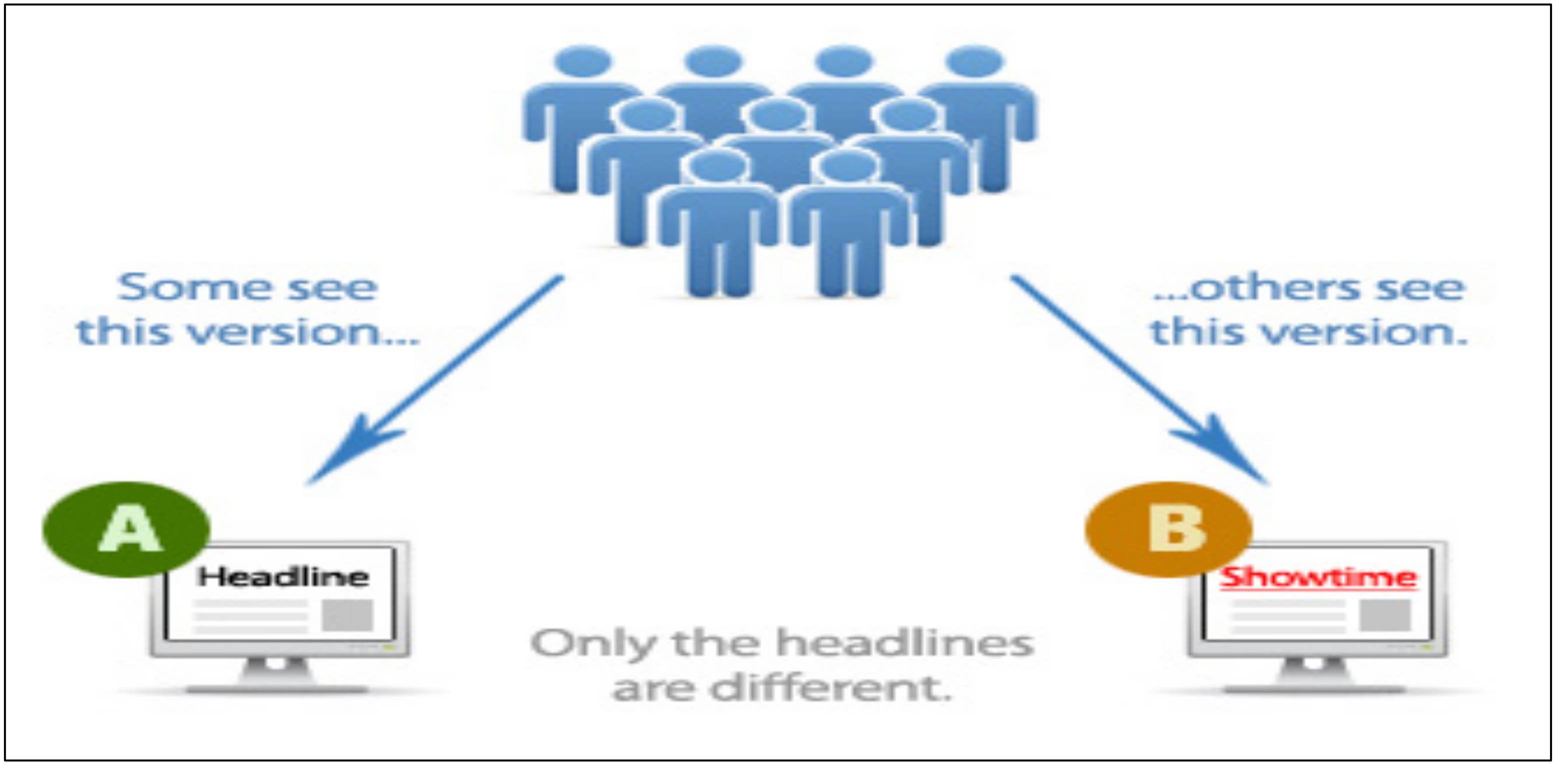
- **A 3^{3-1} fractional factorial design can be constructed from the Latin square as follows:**

3-LEVEL *FRACTIONAL* FACTORIAL DESIGNS

- **A 3^{4-2} fractional factorial design can be constructed from the Graeco-Latin square as follows:**

ON A/B TESTING …

“A/B TESTING (SPLIT TESTING)”



Result
A: 50 signups, B: 75 signups → Version B is better.

AND HAS BEEN AROUND FOR A WHILE

A/B TESTING IS HAPPENING ALL THE TIME

- **Typically runs continuously, often with multiple versions (A/B/C/D) – Facebook, Google, Amazon, Expedia etc. all openly acknowledge this**
- **Variety of new software packages being introduced to make this easy to do (e.g. "Max A/B", "Google Analytics", "Content Experiments" etc.)**
- **One operational note is to think about test power – that is, having enough sample to test differences in <u>subgroups</u>**

MORE DETAILS

- You have two designs of a website: A and B. Typically, A is the existing design (called the control), and B is the new design.
- You split your website traffic between these two versions and measure their performance using metrics that you care about such as conversion rate, sales, bounce rate, etc.
- In the end, you select the version that performs best.
- When doing A/B testing, never wait to test the variation until after you’ve tested the control. Always test both versions simultaneously.

BE CAREFUL!

- **A common mistake when implementing A/B testing is when sequential A/B tests are performed in an effort to arrive at optimal levels for multiple factors.**
- ***A/B tests assess <u>one level of one factor</u> versus the control group, but <u>cannot measure the interaction</u> effect across factors.***
- **By not being able to capture interaction effects, this sequential approach <u>may miss the optimum</u> altogether.**
- **It is more appropriate to perform a multivariate test (factorial test to be exact), where all factors are changed together and all combinations are accounted for.**

MATH BEHIND A/B TESTING

95% Confidence Interval for the rate *p: p % ± 2 * SE*

Standard Error (SE) = Square root of (p*(1-p) /n)

Variation	Conversions / Views	Conversion Rate	Change	Confidence
Variation A (Control)	320 / 1064	30.08% ±2.32%	—	—
Variation B	*250 / 1043*	*23.97% ±2.18%*	*-20.30%*	✓ *99.92%*

- The percentage change of the conversion rate between the Test variation and the Control variation:

$$ChangePercent = \frac{p_{variationB} - p_{variationA}}{p_{variationA}}$$

Confidence: this column reports the significance, or how different the confidence interval for the conversion rate for the Test variation is when compared to the Control variation (this must be at least 95% confident before being flagged as significant)

$$ZScore = \frac{p_{variationB} - p_{variationA}}{\sqrt{SE^2_{variationA} + SE^2_{variationB}}}$$

→ -3.17 with left tail area 0.0008 (=p-value)

→ Confidence = 1 minus p-value.

A/B TESTING – ONLY ONE VARIABLE (FACTOR)

Advantages:

- Quick and easy method.
- Understanding the data is simple, and does not require a lot of analyzing.

- Limitations:
  - **A/B testing does not guarantee the best solution for design and copy.**
  - Does not reveal any information about interaction between variables.

MULTIVARIATE TESTING (MORE THAN ONE FACTOR)

Advantages:

- Can test many different elements on the page in a variety of ways.
- Can find not only the most successful design, but also reveal which elements have the greatest positive or negative impact on a visitor's interaction.

- Limitations:
  - Since all experiments are fully factorial, it requires high traffic to complete the test.
  - Results are fairly difficult to interpret.

OBAMA 2008 CAMPAIGN

Dan Siroker is the co-founder and CEO of Optimizely.

He was Director of Analytics for the Obama 2008 campaign.

***Optimizely* is an optimization platform that provides A/B testing, multivariate testing, and personalization for websites and mobile applications.**

Obama 2008 Campaign

OBAMA 2008 CAMPAIGN (CONT.)

The first experiment was conducted in December 2007.

This experiment helped to raise additional

$60 million in donations.

The experiment tested two parts of the splash page of the campaign's website : the ‘Media’ section at the top and the call-to-action ‘Button’.

OBAMA 2008 CAMPAIGN (CONT.)

They used Google Website Optimizer and ran this experiment as a full-factorial multivariate test.

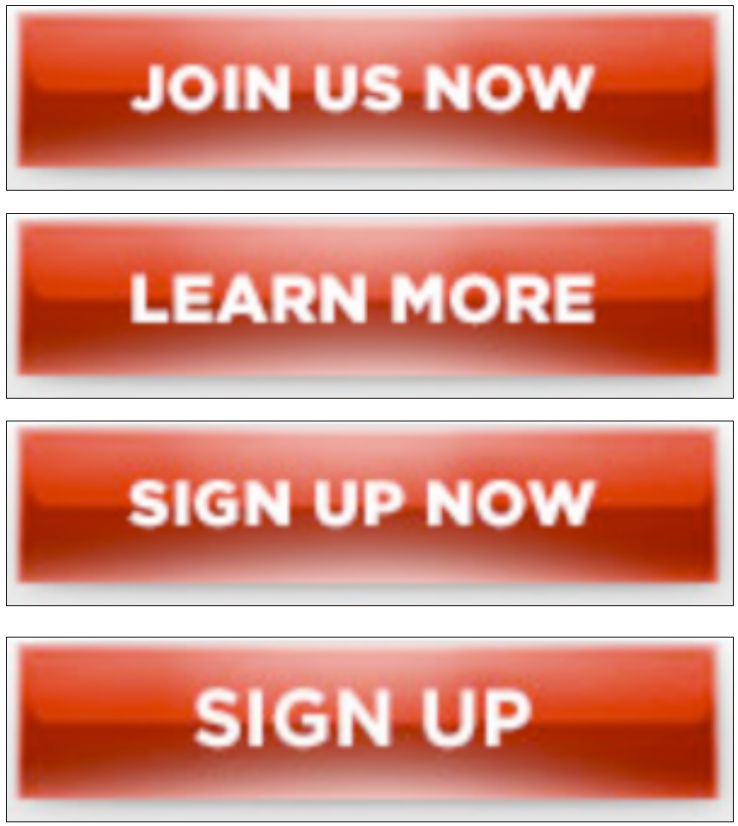
Factors:

- Button (4 levels)
- Media (6 levels: three images and three videos)

24 combinations (4 * 6):

Every visitor was randomly shown one of the combinations.

OBAMA’08 –
BUTTON VARIATIONS



OBAMA'08 – RESULTS

Sign-up rate was used to measure success of the combinations.

Sign-Up Rate = the number of people who signed up/

the number of people who saw that particular variation

Total number of visitors = 310,382 (approximately 13,000 per variation)

OBAMA'08 – RESULTS (CONT.)

Sign-up rates for the combinations of the different sections shown below:

Combinations (24) | Page Sections (2) — Download: XML CSV TSV | Print

Disable | All Combinations (24) | Key: Winner Inconclusive Loser

Combination	Status	Est. conv. rate	Chance to Beat Orig.	Observed Improvement	Conv./Visitors
Original	Enabled	8.26% ± 0.5%	—	—	1088 / 13167
Top high-confidence winners. Run a follow-up experiment »					
Combination 11	Enabled	11.6% ± 0.6%	100%	40.6%	1504 / 12947
Combination 7	Enabled	10.3% ± 0.6%	100%	24.0%	1340 / 13073
Combination 3	Enabled	9.80% ± 0.6%	99.7%	18.7%	1277 / 13025
Combination 10	Enabled	9.23% ± 0.6%	95.9%	11.7%	1203 / 13031
Combination 8	Enabled	9.03% ± 0.6%	91.6%	9.28%	1178 / 13046
Combination 9	Enabled	8.77% ± 0.6%	81.8%	6.10%	1111 / 12672
Combination 6	Enabled	8.64% ± 0.5%	75.3%	4.58%	1108 / 12822

OBAMA'08 – THE WINNER

OBAMA'08 – BENEFITS OF THE EXPERIMENT

	The Original Page	The Winning Variation	Difference
Sign-up Rate	8.26%	11.6%	40.6% improvement
Signed Up	7,120,000 people	10,000,00 people	2,880,000 email addresses
Number of volunteers	712,000	1,000,000	288,000
$21 donation per email address	$149,520,000	$210,000,000	> $60,000,000