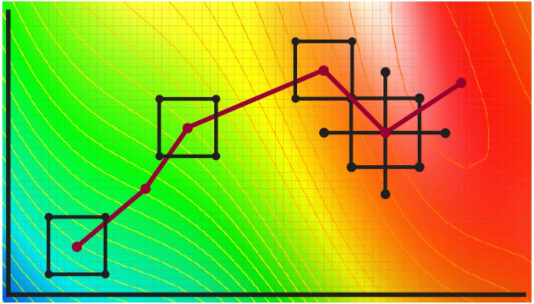
Experimentation for Improvement



http://learnche.org/

Design and Analysis of Experiments

Copyright, sharing, and attribution notice

		Measurable (quantify the variable)	
		No	**Yes**
Controllable	**No**	Disturbances	Covariates
	Yes		Factors

Can you classify all the variables in your experimental system?

		Measurable (quantify the variable)	
		No	**Yes**
Controllable	**No**	Disturbances	Covariates
	Yes		Factors

They should fall into one of these 3 categories.

The “CalApp” example continues

	Low level $-$	**High level** $+$
A: “Promotion”	1 free in-app upgrade	30-day trial of all features
B: “Message”	“CalApp has your schedule available at your fingertips, on any device.”	“CalApp features are configurable; only pay for the features you want.”
C: “Price”	in-app purchase price is 89c	in-app purchase price is 99c

A subtlety: characteristics of a nuisance factor

[Flickr: nickwebb]

- the factor does vary during your experiments
- it is controllable, and it is measurable (just like a regular factor)
- the factor is not the focus of your experiment

How do we deal with a nuisance factor to avoid bias?

- randomization will minimize the effect, but not eliminate it
- your experiments should be successful, despite the nuisance factor
- we use a process called blocking

"Does the process have to work successfully with different levels of the nuisance variable?"

If **yes**: you must actively plan for blocking (that's the next topic)

If **no**: that indicates you have good control over your system

Some examples of nuisance factors

CalApp

Flickr: williamhook

Apple or Android

Baking

Brand of flour

Shift work

Flickr: leehaywood

Day shift or night shift

Gas mileage

Driver 1 or driver 2

The “CalApp” example continues

	Low level $-$	**High level** $+$
A: “Promotion”	1 free in-app upgrade	30-day trial of all features
B: “Message”	“CalApp has your schedule available at your fingertips, on any device.”	“CalApp features are configurable; only pay for the features you want.”
C: “Price”	in-app purchase price is 89c	in-app purchase price is 99c

Planning for blocking: add a new factor column to the standard order table

Run	**A** "Promotion"	**B** "Message"	**C** "Price"	**D** "Apple or Android"
1	$-$	$-$	$-$	
2	$+$	$-$	$-$	
3	$-$	$+$	$-$	
4	$+$	$+$	$-$	
5	$-$	$-$	$+$	
6	$+$	$-$	$+$	
7	$-$	$+$	$+$	
8	$+$	$+$	$+$	

- Factor **D**: experiment with either Android $(-)$ or Apple $(+)$ users
- This would require 16 experiments for a full factorial in factors **A**, **B**, **C**, and **D**
- but, we only want 8 experiments

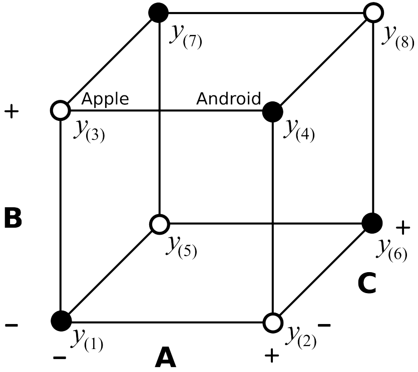
Planning for blocking: add a new factor column to the standard order table

Run	**A** "Promotion"	**B** "Message"	**C** "Price"	**D=ABC** "Apple or Android"
1	$-$	$-$	$-$	$-$ Android
2	$+$	$-$	$-$	$+$ Apple
3	$-$	$+$	$-$	$+$ Apple
4	$+$	$+$	$-$	$-$ Android
5	$-$	$-$	$+$	$+$ Apple
6	$+$	$-$	$+$	$-$ Android
7	$-$	$+$	$+$	$-$ Android
8	$+$	$+$	$+$	$+$ Apple

- Factor **D**: experiment with either Android $(-)$ or Apple $(+)$ users
- This would require 16 experiments for a full factorial in factors **A**, **B**, **C**, and **D**
- but, we only want 8 experiments

Visualizing the blocking procedure on a cube plot

Run	**A** "Promotion"	**B** "Message"	**C** "Price"	**D=ABC** "Apple or Android"
1	−	−	−	− Android
2	+	−	−	+ Apple
3	−	+	−	+ Apple
4	+	+	−	− Android
5	−	−	+	+ Apple
6	+	−	+	− Android
7	−	+	+	− Android
8	+	+	+	+ Apple



Planning for blocking: add a new factor column to the standard order table

Run	A "Promotion"	B "Message"	C "Price"	D = ABC "Apple or Android"	Outcome, y
1	$-$	$-$	$-$	$-$ Android	$y_{(1)}$
2	$+$	$-$	$-$	$+$ Apple	$y_{(2)}$
3	$-$	$+$	$-$	$+$ Apple	$y_{(3)}$
4	$+$	$+$	$-$	$-$ Android	$y_{(4)}$
5	$-$	$-$	$+$	$+$ Apple	$y_{(5)}$
6	$+$	$-$	$+$	$-$ Android	$y_{(6)}$
7	$-$	$+$	$+$	$-$ Android	$y_{(7)}$
8	$+$	$+$	$+$	$+$ Apple	$y_{(8)}$

- Android users: $y_{(i)} + g$
- Apple users: $y_{(i)} + h$

Planning for blocking: add a new factor column to the standard order table

Run	A "Promotion"	B "Message"	C "Price"	D = ABC "Apple or Android"	Outcome, y
1	$-$	$-$	$-$	$-$ Android	$y_{(1)} + g = \widetilde{y}_{(1)}$
2	$+$	$-$	$-$	$+$ Apple	$y_{(2)} + h = \mathring{y}_{(2)}$
3	$-$	$+$	$-$	$+$ Apple	$y_{(3)} + h = \mathring{y}_{(3)}$
4	$+$	$+$	$-$	$-$ Android	$y_{(4)} + g = \widetilde{y}_{(4)}$
5	$-$	$-$	$+$	$+$ Apple	$y_{(5)} + h = \mathring{y}_{(5)}$
6	$+$	$-$	$+$	$-$ Android	$y_{(6)} + g = \widetilde{y}_{(6)}$
7	$-$	$+$	$+$	$-$ Android	$y_{(7)} + g = \widetilde{y}_{(7)}$
8	$+$	$+$	$+$	$+$ Apple	$y_{(8)} + h = \mathring{y}_{(8)}$

- Android users: $\widetilde{y}_{(i)} = y_{(i)} + g$
- Apple users: $\mathring{y}_{(i)} = y_{(i)} + h$

Why does blocking like this work so successfully? (math alert!)

The main effect of **A**

$$\mathbf{A} = \frac{1}{2}\left[\frac{\left(\mathring{y}_{(8)} - \widetilde{y}_{(7)}\right) + \left(\widetilde{y}_{(4)} - \mathring{y}_{(3)}\right) + \left(\widetilde{y}_{(6)} - \mathring{y}_{(5)}\right) + \left(\mathring{y}_{(2)} - \widetilde{y}_{(1)}\right)}{4}\right]$$

Notice two $+\widetilde{y}$ values and two $-\widetilde{y}$; and also two $+\mathring{y}$ values and two $-\mathring{y}$

$$\mathbf{A} = \frac{1}{8}\Big[\left(y_{(8)} + h - y_{(7)} - g\right) + \left(y_{(4)} + g - y_{(3)} - h\right) + \left(y_{(6)} + g - y_{(5)} - h\right) + \left(y_{(2)} + h - y_{(1)} - g\right)\Big]$$

which simplifies to

$$\mathbf{A} = \frac{1}{8}\Big[-y_{(1)} + y_{(2)} - y_{(3)} + y_{(4)} - y_{(5)} + y_{(6)} - y_{(7)} + y_{(8)} \cancelto{0}{-2g + 2g} \cancelto{0}{-2h + 2h}\Big]$$

$\mathbf{A}$ = pure effect of **A**, without bias

Why does blocking like this work so successfully? (math alert!)

In a similar way, you can show all effects are estimated without bias:

- **A**
- **B**
- **C**
- **AB**
- **AC**
- **BC**

- except, for the effect of **ABC**:

$$\mathbf{ABC} = \frac{1}{8}\Big[\underbrace{-y_{(1)} + y_{(2)} + y_{(3)} - y_{(4)} + y_{(5)} - y_{(6)} - y_{(7)} + y_{(8)}}_{\text{pure effect of } \textbf{ABC}} \underbrace{-4g + 4h}_{\textit{with} \text{ bias}}\Big]$$

Why does blocking like this work so successfully? (math alert!)

In a similar way, you can show all effects are estimated without bias:

- **A**
- **B**
- **C**
- **AB**
- **AC**
- **BC**

- except, for the effect of **ABC**:

$$\mathbf{ABC} = \frac{1}{8}\Big[\underbrace{-y_{(1)} + y_{(2)} + y_{(3)} - y_{(4)} + y_{(5)} - y_{(6)} - y_{(7)} + y_{(8)}}_{\text{pure effect of } \textbf{ABC}} \underbrace{\color{red}{-4g\ +4h}}_{\text{confounded with the blocking effect}}\Big]$$

More than two blocks are easily possible

CalApp

Flickr: williamhook

Apple, Android, Blackberry

Baking

4 different brands of flour

Shift work

Flickr: leehaywood

Day shift
Afternoon shift
Night shift

Gas mileage

Gasoline brand 1
Gasoline brand 2
Gasoline brand 3

TABLE 12.14. Construction and blocking of some designs of resolution V and higher so that no main effect or interaction is confounded with any other main effect or interaction

(1) number of variables	(2) number of runs	(3) degree of fractionation	(4) type of design	(5) method of introducing "new" factors	(6) blocking (with no main effect or interaction confounded)	(7) method of introducing blocks
5	16	$\frac{1}{2}$	2_{V}^{5-1}	$\pm 5 = 1234$	not available	
6	32	$\frac{1}{2}$	2_{VI}^{6-1}	$\pm 6 = 12345$	two blocks	$B_1 = 123$
	64	1				

General rule for blocking in two blocks

1. Write out your existing standard order table
2. Add an extra "factor" to your standard order table
3. Generate a half-fraction, using the trade-off table, on this new factor
4. Signs with $-$ are in one block; and signs with $+$ are in the other block

Run	A	B	C	D=ABC
1	$-$	$-$	$-$	$-$ Android
2	$+$	$-$	$-$	$+$ Apple
3	$-$	$+$	$-$	$+$ Apple
4	$+$	$+$	$-$	$-$ Android
5	$-$	$-$	$+$	$+$ Apple
6	$+$	$-$	$+$	$-$ Android
7	$-$	$+$	$+$	$-$ Android
8	$+$	$+$	$+$	$+$ Apple