Design Of Experiments (DOE)

A CMER Science Presentation

Link to adaptive management

Researchers often discover after running an experiment that the data are insufficient to meet the objectives.

Experiment

A step in the scientific method that arbitrates between competing models or hypotheses.

Statistics helps make effective use of experimental data and includes guidance for data collection.

Experimental approaches

- Build-Test-Fix
- Evaluate One Factor At A Time (OFAAT)
- Designed Experiments
  - Screening
  - Factorial (2^k, full, fractional)
  - Augment
  - Others…

Build – Test – Fix

The way most of us make decisions.

- Do something
- Perhaps notice a relationship between what we did and some outcome
- Perhaps alter the process
- Repeat.

– Requires intuition and luck.

– Unlikely to yield optimal results.

One Factor At A Time

The way most of us do experiments:

1. Identify a key factor.
2. Make changes to the key factor while trying to hold all other factors constant.
3. Evaluate the effect.
4. Alter the process (maybe).
5. Repeat.

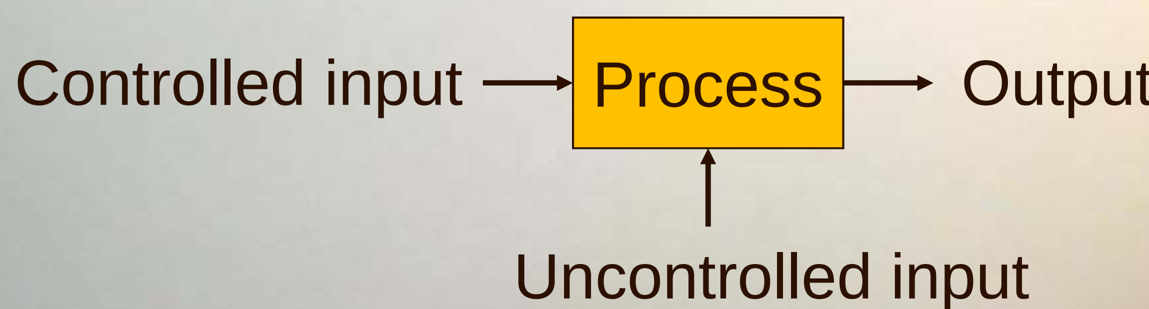
OFAAT analysis

- Contingency table
- Bivariate (regression)
- Oneway ANOVA (or t-test)

Good when you only have one factor of interest, but slow and inefficient and likely to miss interactions among multiple factors.

Design Of Experiments

DOE is the purposeful control of inputs (factors) in such a way as to deduce their relationships (if any) with the output (responses)



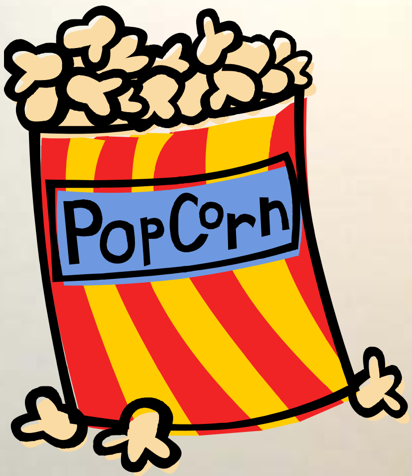
Design Of Experiments

Methodology to achieve predictive knowledge from a complex multi-variable process with the fewest trials possible

- Screening
- Modeling
- Optimization

Popcorn example

Problem statement: How to get the most from a bag of microwave popcorn?



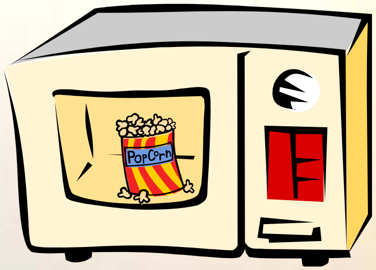
Response

Number of edible kernels

- Alternate: Proportion of kernels that are edible.

Input factors

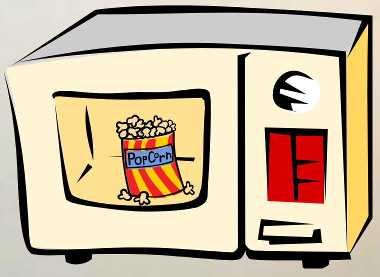
- Brand of microwave popcorn
- ~~Color of the bag~~
- ~~Weight of the bag~~
- ~~Type of microwave oven~~
- Power setting
- Time setting



Key step: Understand the process!

Would you expect there to be an interaction between time and power?

i.e., if a time 5 minutes produces the most popcorn at a power setting of 5, will 5 minutes still produce the most popcorn edible popcorn at a power setting of 10?



Importance

No interaction

Power (rows) × Time (columns)

Power	High / Low	High / High
	Low / Low	Low / High

Time

Interaction

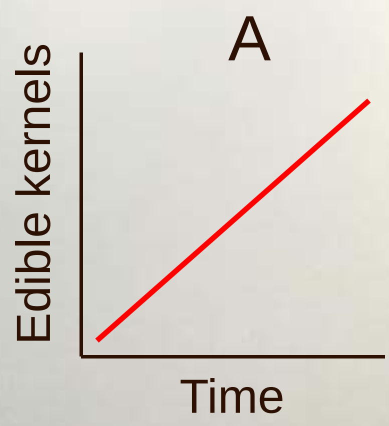
Power	High / Low	High / Med	High / High
	Med / Low	Med / Med	Med / High
	Low / Low	Low / Med	Low / High

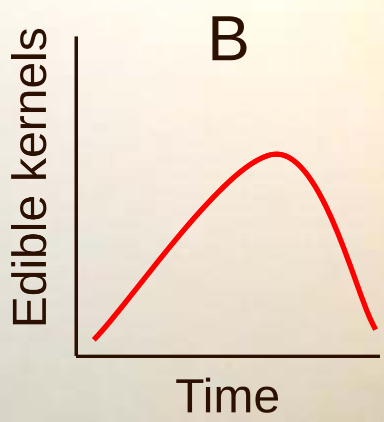
Time

Key step: Understand the process!

Are the effects of time and power likely to be linear or non-linear?

i.e., at a power setting of 10, is the number of edible kernels likely to look like A or B?

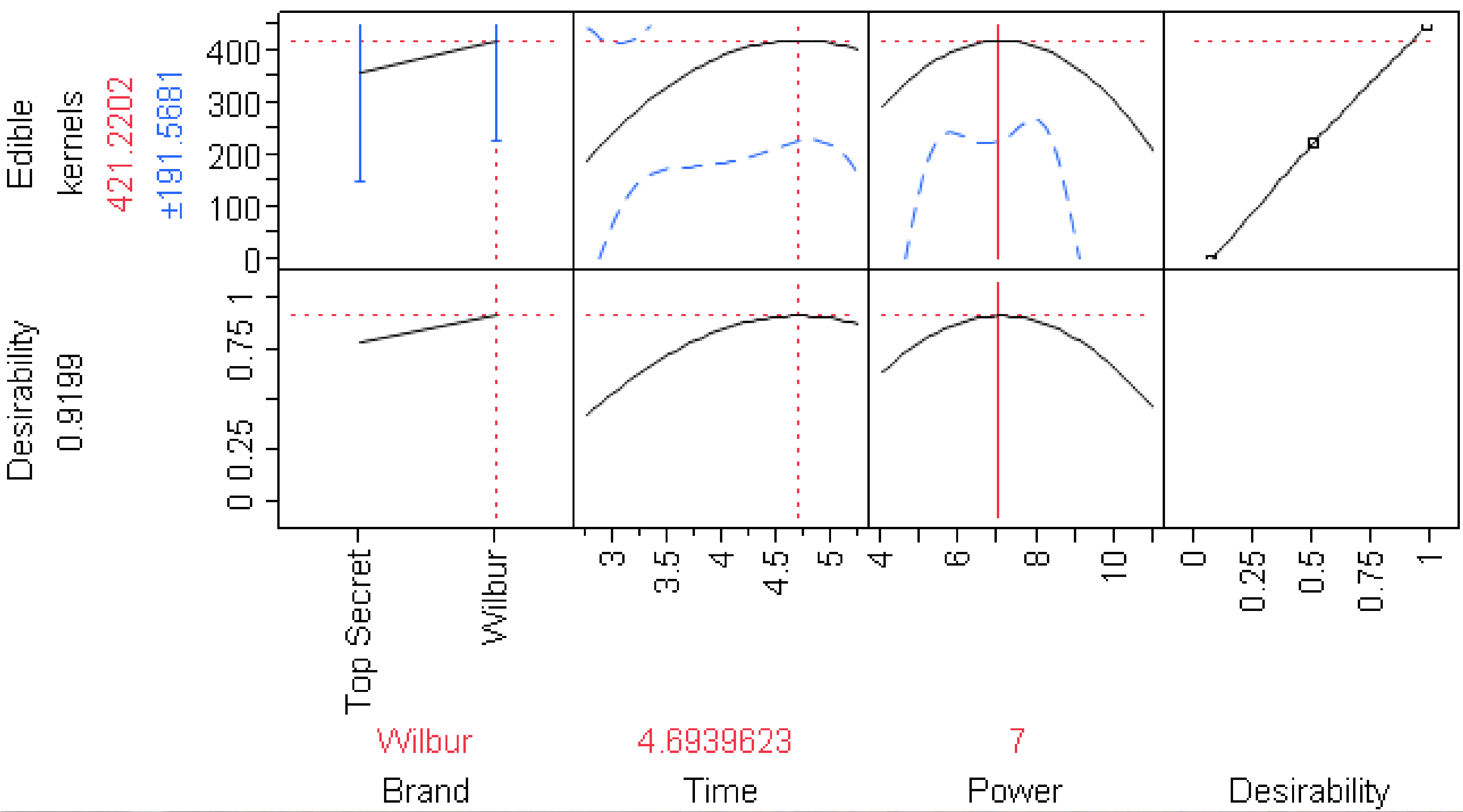




DOE:

Group	Brand	Time	Power	Edible kernels
1	Top Secret	3	7	30
2	Top Secret	3	7	120
3	Top Secret	3	10	120
4	Top Secret	4	8	250
5	Top Secret	5	5	370
6	Top Secret	5	5	400
7	Top Secret	5	8	350
8	Top Secret	5	8	370
9	Wilbur	3	7	374
10	Wilbur	3	8	340
11	Wilbur	3	10	20
12	Wilbur	4	6	440
13	Wilbur	4	9	340
14	Wilbur	5	5	170
15	Wilbur	5	7	370
16	Wilbur	5	8	420

Optimized output



Process understanding is key!

Statistical methods do not prove that a factor (or factors) has a particular effect. They only provide guidelines to the reliability and validity of results.

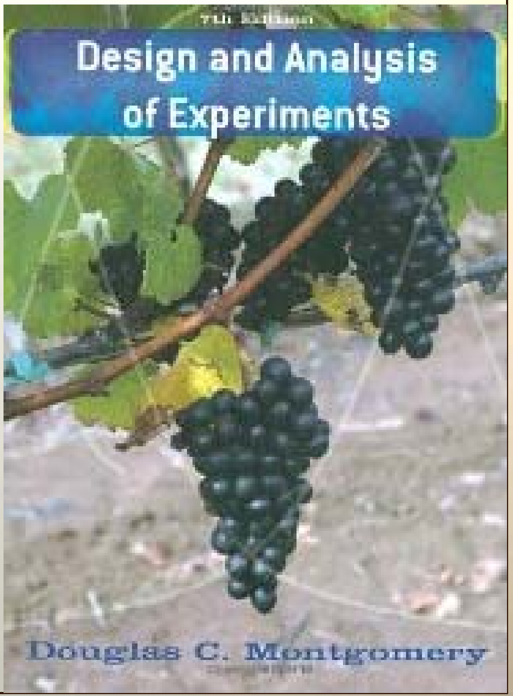
Sound conclusions are developed by combining statistics with good process knowledge.

Process understanding is key!

How many of us are statisticians?

References

Design and analysis of experiments by Douglas C. Montgomery



Lots more.

Genesis of this talk

Accuracy and Bias study

- Selection of FPA's
  - Completely random?
  - Stratified random?
  - How about purposeful with a random element?

How would we design the study **if** we could design the FPA as part of an experiment (e.g., natural experiment)?

DOE steps

1. Problem statement
2. Choice of response variable(s)
3. Choice of factors, levels, and ranges
4. Choice of experimental design
5. Performing the experiment
6. Statistical analysis
7. Conclusions and recommendations

Important considerations

Randomization!

- Independence
- Even out uncontrollable factors

Replication

- Error estimation
- Accuracy

Blocking

- Control for known 'nuisance' factors
- Precision

Selecting factor levels

- The bigger a factors effect, the easier it is to detect.
- If factor's levels are close, measured effect may be statistically insignificant
- Spanning entire region of interest is likely to yield the most understanding.

Full factorial designs

- Easy design to create. Each factor is tested at each condition of the factor
- Popcorn example: 2 brands x 6 power settings x 3(?) time settings = 36 trials
- Results analyzed with a linear model
- Cost: resources, time, materials. What if you can't afford 36 trials? What if you only have 16 bags of popcorn?

2^k Factorial designs

- Test each of k factors at 2 levels
- Great for screening (to find important factors)
- Cannot identify nonlinear effects
- Follow-up with additional sampling on important factors (augment design).

Standard 2^3 Layout

Run	A	B	C	Response
1	-	-	-	
2	-	-	+	
3	-	+	-	
4	-	+	+	
5	+	-	-	
6	+	-	+	
7	+	+	-	
8	+	+	+	

2^3 Layout (with results)

Run	Temp	Time	Ingredient	% Yield
1	200	1	1	65.3
2	200	1	2	81.3
3	200	2	1	53.3
4	200	2	2	69.9
5	250	1	1	61.8
6	250	1	2	77.4
7	250	2	1	73.9
8	250	2	2	89.9

Randomization

Although we write our test matrix in standard order, we should actually perform the 8 experiments in RANDOM order

- Randomization will make any factor we overlooked likely to contribute to random uncertainty rather than systematic error

Fractional Factorial

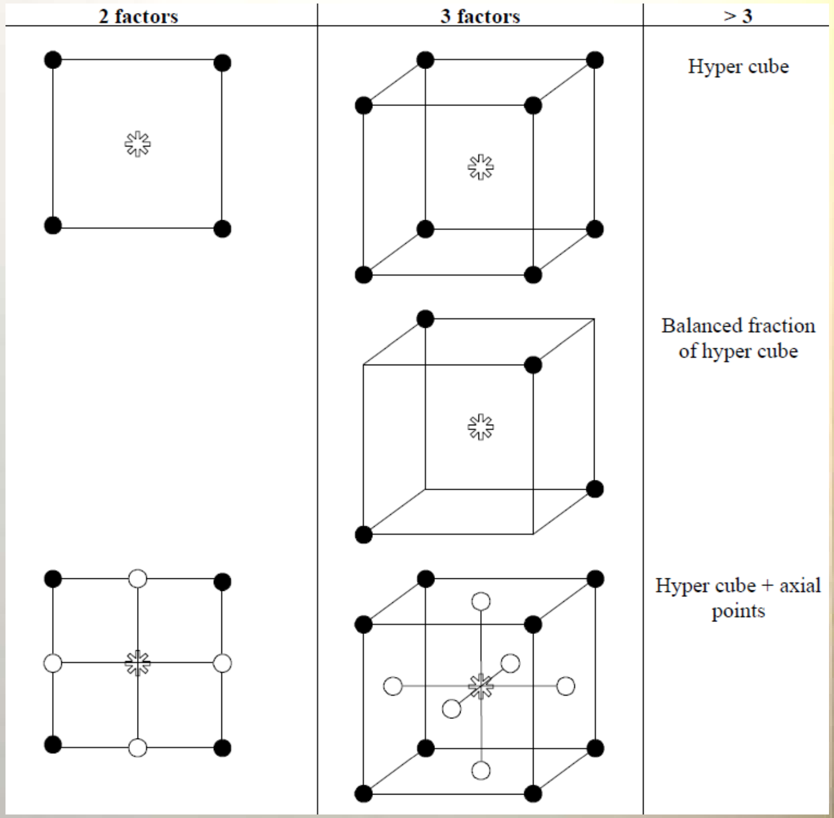
- To do a full 2^k factorial experiment, you must do at least 2^k runs
  - Testing 4 factors costs twice as much as testing 3 factors.
- You can test more factors in the same 8 runs if you settle for a “fractional” factorial design

Fractional Factorial

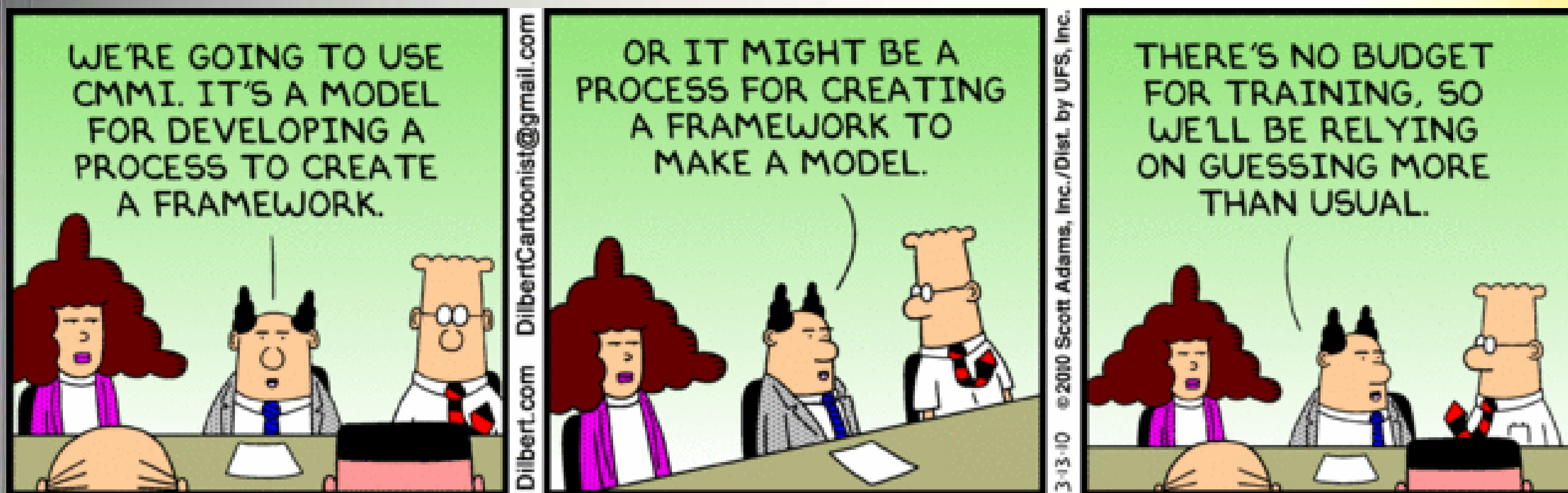
- Less than full factorial
- Condition combinations are chosen to provide sufficient information to determine the factor effect
- Key point, what subset of factor combinations?
  - Pick the wrong set and you're likely to miss interactions

Don’t leave it up to fate

Fractional factorial designs require specific combinations of factors.



DOE software can help



Data accuracy

Garbage in - Garbage out (or worse)

Bad data typically leads to days, months, and perhaps years of wasted effort trying to interpret the bad data. More effort is wasted in follow-up studies.

Pseudo replication

- Beware of spatial or temporal pseudo replication
  - Repeated measurements from the same individual (time series).
  - Spatial involves several measurements from the same vicinity (nested or split plot designs).

Tangent: Data mining

- Process of extracting patterns from data.
  - Neural networks
  - Recursive partitioning
- Used on very large datasets
- Good for developing hypotheses

Data mining myths*

Data mining will (not)

- find answers to unasked questions
- eliminate the need to understand the process
- eliminate the need to collect good data
- eliminate the need to have good data analysis skills

*Excerpted from a talk by Richard De Veaux at the Exploring Interactive and Visual Data Mining with JMP seminar in Seattle on January 21, 2010

Data mining

Take home message from Dick De Veaux

- Data mining gives clues that may be helpful in generating hypotheses
- Follow-up with DOE

Conclusions

- DOE helps define
  - The main and interaction effects in the process?
  - What settings would bring about less variation in the output?
  - What settings would the process deliver optimal performance?

Resource Allocation

- Don't commit all resources to one design
  - Start with Screening design
  - Only 25% of resources on any one experiment
- Learn from each design
  - What did you do wrong?
    - Excluded factors, wrong conditions, etc.
  - What to do next?
    - Sometimes next stage of improvement isn't worth the cost of another experiment

Selecting factors

- What are all the control factors that affect the response?
  - For each response, brainstorm likely factors
  - For screening, if more than 5-7 factors:
    - Reduce factor list through ranking (e.g., Nominal Group technique)
    - Hold some factors constant (e.g., FPA type)

Selecting factors

- Over what range does it make sense to change these factors?
- Are any combinations of factors likely to interact, and are any not impossible to achieve?

Don't forget!

- Randomization
- Do we need replication?
- Do we need blocking?

Final comment

Ask good questions. DOE is only as good at answering questions as you are at asking them!

If signal is twice the noise, then life is easy. If signal is half the noise, then life is tough.