Operations Research
and Complex Systems
Department of Electrical Engineering
Universidade Federal de Minas Gerais

UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Design and Analysis of Experiments

Prof. Felipe Campelo, Ph.D.

Dept. Electrical Engineering, Universidade Federal de Minas Gerais

Belo Horizonte, Brazil - September 29 2015

Motivation

Why experiment?

“*What do you think science is?*
There’s nothing magical about science.
It is simply a systematic way for carefully
and thoroughly observing nature and
using consistent logic to evaluate results.”

Steven P. Novella

Image: `http://www.relativelyinteresting.com/definition-science-steven-novella/`

Experiments

Definition of experiment

An experiment can be characterized as a test (or a series of tests) wherein changes are introduced in the state of a system or process, enabling the observation and characterization of effects that can occur as a result of these changes.

Usually performed to:

- Uncover influential variables in a given system or process;
- Compare the performance of different alternatives;
- Characterize the behaviour of a system or process under study.

Image: `http://xkcd.com`

Experiments

Data gathering

- Retrospective study;
- Observational study;
- Designed experiment;

Characteristics

- Use of historical data;
- Investigating correlations;

Drawbacks

- Data representativeness;
- Availability of data;

Image: `http://www.computer.org/web/computingnow/archive/march2013`

Experiments

Data gathering

- Retrospective study;
- **Observational study**;
- Designed experiment;

Characteristics

- Observation of the system with minimal disturbance;
- Investigation of usual behaviors;

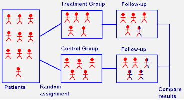
Drawbacks

- Low representativeness of extreme cases;
- Low variability can affect observation of interesting effects;

Image: `http://goo.gl/A51KKM`

Experiments

Data gathering



- Retrospective study;
- Observational study;
- Designed experiment;

Characteristics

- Introduction of deliberate changes in the system;
- Inference on *causal* relationships;

Drawbacks

- Requires rigorous experimental design and data analysis;

Image: `http://library.downstate.edu/EBM2/2200.htm`

Experimentation strategies

Educated guessing

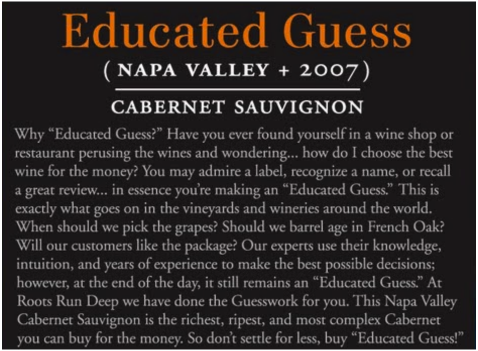
- Select arbitrary combination of values for the experimental factors;
- Test and observe behavior;
- Change one or two values at a time, rinse and repeat;

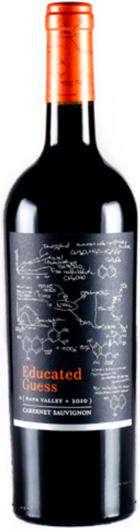
- Widely used in engineering;
- Can achieve good results, but has a lot of limitations;

Experimentation strategies

Educated guessing

- Select arbitrary combination;
- Test an observe;
- Change and re-test;



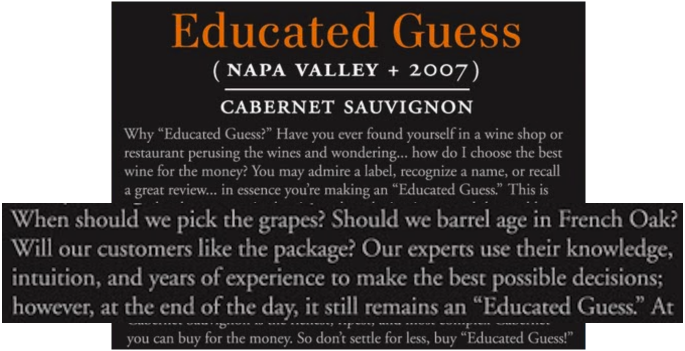


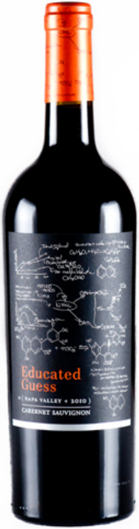
Images (c) Roots Run Deep Winery:

Experimentation strategies

Educated guessing

- Select arbitrary combination;
- Test an observe;
- Change and re-test;





Images (c) Roots Run Deep Winery:

Experimentation strategies

COST: Change One Separate factor at a Time

- Select a reference point;
- Change each factor individually, keeping all others constant;

- Widely used in practice;
- Can achieve good results as long as there are no interaction effects;

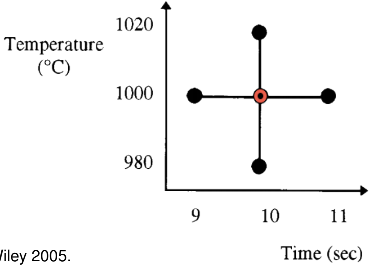


Image: D.C. Montgomery, *Design and Analysis of Experiments*, Wiley 2005.

Experimentation strategies

Factorial designs

- Select **levels** for each factor;
- Change levels simultaneously, in a systematic way;

- Estimation of main effects and interactions;
- Greater precision in the effect estimates;

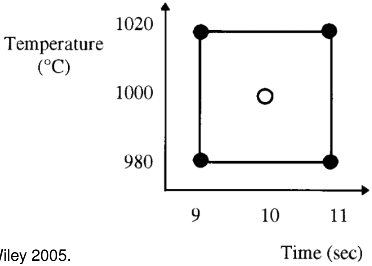


Image: D.C. Montgomery, *Design and Analysis of Experiments*, Wiley 2005.

Pre-experimental checklist

Before you even start...

Is the comparison relevant?

Will the (possible) results be of interest to anyone?

Does it have any practical implications?

How does it fit within the literature?

Pre-experimental checklist

Before you even start...

Is the comparison relevant?

Will the (possible) results be of interest to anyone?

Does it have any practical implications?

How does it fit within the literature?

However...

"*Sometimes one should do a completely wild experiment,*
like blowing the trumpet to the tulips every morning
for a month. Probably nothing would happen, but what if it did?"

– Sir George Howard Darwin

Girl playing: `http://www.film.queensu.ca/tulips/default.html`

Pre-experimental design

The quest for *The Question*

"- Would you tell me, please, which way I ought to go from here?
- That depends a good deal on where you want to get to, said the Cat.
- I don't much care where - said Alice.
- Then it doesn't matter which way you go, the Cat replied

Lewis Carroll, **Alice in Wonderland**

The first (and possibly the most important) thing to determine is what exactly your experiment is intended to reveal / discover.

Image adapted from `https://www.pinterest.com/pin/381609768395834338/`

Pre-experimental design

The quest for *The Question*

What is the purpose of your experiment?

> “*I want to prove that my proposed method is good!*”

Lets drop the exclamation mark and refine that a little, shall we?

What is the purpose of your experiment?

> "*I want to prove that my proposed method is good!*"

Lets drop the exclamation mark and refine that a little, shall we?

> "*I want to **discover if** my proposed method is good.*"

That's much better.

Pre-experimental design

The quest for *The Question*

What is the purpose of your experiment?

> “*I want to prove that my proposed method is good!*”

Lets drop the exclamation mark and refine that a little, shall we?

> “*I want to **discover if** my proposed method is good.*”

That’s much better. Now, “good” compared to what?

> “*I want to discover if my proposed method is **better than method A**.*”

We’re making progress...

Pre-experimental design

The quest for *The Question*

Lets proceed with our quest for *The Question*: you want to compare your method to another, but on what exactly?

Pre-experimental design

The quest for *The Question*

Lets proceed with our quest for *The Question*: you want to compare your method to another, but on what exactly?

> "*I want to discover if my proposed method is better than method A **in terms of X**.*"

Pre-experimental design

The quest for *The Question*

Lets proceed with our quest for *The Question*: you want to compare your method to another, but on what exactly?

> "*I want to discover if my proposed method is better than method A* ***in terms of X***."

And how are you going to measure X?

Pre-experimental design

The quest for *The Question*

Lets proceed with our quest for *The Question*: you want to compare your method to another, but on what exactly?

> "*I want to discover if my proposed method is better than method A* ***in terms of X***."

And how are you going to measure X?

> "*I want to discover if my proposed method is better than method A in terms of X,* ***measured using indicator*** $\mathcal{F}$."

Pre-experimental design

The quest for *The Question*

Lets keep it up just a little more, we're almost there. When you say "better", what exactly do you mean? Typical, best, worst case?

Pre-experimental design

The quest for *The Question*

Lets keep it up just a little more, we're almost there. When you say "better", what exactly do you mean? Typical, best, worst case?

> "*I want to discover if my proposed method is better than method A in terms of X, measured using **the mean of** indicator $\mathcal{F}$.*"

Lets keep it up just a little more, we're almost there. When you say "better", what exactly do you mean? Typical, best, worst case?

> "*I want to discover if my proposed method is better than method A in terms of X, measured using* ***the mean of*** *indicator $\mathcal{F}$.*"

And you want to investigate if your method is good for what?

Pre-experimental design

The quest for *The Question*

Lets keep it up just a little more, we're almost there. When you say "better", what exactly do you mean? Typical, best, worst case?

> "*I want to discover if my proposed method is better than method A in terms of X, measured using* ***the mean of*** *indicator* $\mathcal{F}$."

And you want to investigate if your method is good for what?

> "*I want to discover if my proposed method is better than method A, in terms of X (measured using the mean of indicator* $\mathcal{F}$*),* ***for solving problem Y of population Q***."

Pre-experimental design

The quest for *The Question*

The process of determining *The Question* is an important one, which is often overlooked in many fields of engineering.

Besides escaping the mockery of the Cheshire Cat, there are good reasons not to ignore this step:

- *HARKing* tends to greatly increase the rate of false positives in favor of the “proposed approach”;
- Thinking about *The Question* forces the experimenter to consider important aspects of his or her research, such as scope and performance measurement;

Harking = *Hypothesizing After the Results are Known*

Fundamental principles

Design of experiments (DoE)

Process of designing data gathering protocols to enable accurate analyses by statistical tools, capable of supporting sound and objective conclusions.

- Applicable to systems and processes subject to noise, experimental errors, uncertainties, etc.
- Necessary for conclusions to have a quantifiable meaning;
- Helpful in avoiding errors due to personal biases or other artifacts of experimentation and analysis.

Fundamental principles

Design of experiments (DoE)

Design

- Scientific/technical question of interest;
- Experimental variables and values;
- Statistical parameters;
- Protocols for data gathering;

Analysis

- Test statistic;
- Model validation;
- Effect sizes;
- Conclusions and recommendations;

Fundamental principles

Design of experiments (DoE)

- Repetition and replication;
- Randomization;
- Blocking.

- Repeated measurements - estimation of within-group variability;
- Replication - estimative of the experimental error;
- Greater precision in estimating the model parameters;
- Be careful with *pseudoreplication*!

Design of Experiments

Sample size, replication and pseudoreplication - a short detour

Suppose that we want to investigate the question: “*Is the average hair length different between students and professors working in Speech Sciences?*”

Lets assume that the audience of this tutorial is a representative sample of our population of interest;

If we take 5 professors and 5 students from the audience and measure 1 hair from each head, what is our sample size?

Example adapted from B. Shipley, *Cause and Correlation in Biology*. Cambridge University Press, 2000.

Cousin Itt: `http://kawiku.deviantart.com/art/Cousin-Itt-334962933`

Design of Experiments

Sample size, replication and pseudoreplication - a short detour

Suppose that we want to investigate the question: "*Is the average hair length different between students and professors working in Speech Sciences?*"

Lets assume that the audience of this tutorial is a representative sample of our population of interest;

If we take 5 professors and 5 students from the audience and measure 1 hair from each head, what is our sample size?

What if we take 30 hairs from each head?

Example adapted from B. Shipley, *Cause and Correlation in Biology*. Cambridge University Press, 2000.

Cousin Itt: `http://kawiku.deviantart.com/art/Cousin-Itt-334962933`

Design of Experiments

Sample size, replication and pseudoreplication - a short detour

Sampling more hairs from a given head improves the precision of our estimate for that particular head, but ***it does not increase our effective sample size!***

In this example, performing statistical tests considering the 300 individual measurements (10 heads, 30 hairs per head) as independent values would falsely inflate our degrees of freedom.

This common mistake is called *pseudoreplication*, and results in a much higher rate of false positives in statistical tests.

A simple solution is to use the average hair length per head as the individual data points.

Cousin Itt: `http://kawiku.deviantart.com/art/Cousin-Itt-334962933`

Fundamental principles

Design of experiments (DoE)

- Repetition and replication;
- Randomization;
- Blocking;

- Avoids contamination of the data by order/placement dependent effects;
- *Stratified randomization* is often useful to balance experimental groups;

Fundamental principles

Design of experiments (DoE)

- Repetition and replication;
- Randomization;
- Blocking;

- Isolation of nuisance variables that can be controlled;
- Improvement in the estimation of effects for the factors of interest;
- Reduction of residual variance;
- Reduction in sample size requirements;

Fundamental principles

The role of experimental design

Experimental design is useful for avoiding the influence of spurious factors and personal biases on the results, by performing experiments in a impartial and objective way.

> "*The great tragedy of Science - the slaying of a beautiful hypothesis by an ugly fact.*"
> – Thomas H. Huxley

Image: `http://www.iep.utm.edu/huxley/`

Structure of Experimental Design

Main points

To enable the use of a scientific approach in the design of an experiment, it is important to have on a solid understanding of:

- The field where the experiment is to be conducted;
- The strategy for data collection;
- The way the data should be analyzed (at least qualitatively).

Conduction of the Experiment

Follow the plan!

- Data gathering must be consistent with the design, otherwise the validity of the results may be compromised:
  - No premature stops;
  - *No-peeking rule*[1];
- Use of pilot experiments:
  - Gathering of preliminary information;
  - Practice with the experimental conditions;

[1] Except when planned, of course.

Analysis of the experimental data

A consequence of design

- Analysis techniques are generally relatively simple, but *the devil is in the details*;
- Use of existing statistical tools and frameworks, such as
- Free, versatile, good graphical capabilities, relatively simple (but with one hell of a learning curve);

Analysis of the experimental data

Statistical modeling

- General procedure for testing the experimental hypotheses:
  - Definition of a *null-model* (absence of effects) and of a desired level of significance;
  - Determination of $P(data|null\text{-}model)$;
  - Decision by rejection (or not) of the null hypothesis;
  - Validation of model assumptions;
  - Estimation of the *magnitude* of differences - **practical significance**;

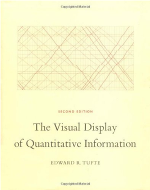
*Statistical methods do not **prove** anything, but they allow an objective definition of margins of plausibility for certain statements.*

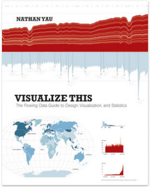
Reporting of results

Presentation

Combine textual, numeric and graphical elements to tell a story with your data. It simplifies the understanding and analysis of the results.

- Strive to achieve graphical excellence;
- Coherence of notation - special attention to figures and tables;
- Display simultaneous confidence intervals and other graphical indicators of effect size.





Other great resources on graphical excellence:

Flowing Data (http://flowingdata.com/)

Information is Beautiful (http://www.informationisbeautiful.net)

Reporting of results

Tell the whole story

Avoid *cherrypicking* your results;

Report and describe anomalous results and outliers (even if they are discarded in the modeling phase);

Exercise *extreme* caution when discarding outliers!

Detail stop criteria, computational cost, and any other relevant information for the understanding and reproducibility of your results.

Whenever possible, share your code! It's good for the field as a whole (for reproducibility), and it is good for your paper (for citations)!

R.D. Peng, *Reproducible Research in Computational Science*, Science 334(6060):1226-1227, 2011

P. Vandewalle, *Code Sharing Is Associated with Research Impact in Image Processing*, Computer Science and Engineering 14(4):42-47, 2012

Conclusions

Drawing and reporting conclusions

Conclusions should be based on solid evidence from the data;

Be conservative - don't exaggerate the generality of the results;

Report significance levels, effect sizes, and the assumptions under which the results are valid;

Suggest explanations to the observed results;

Be careful with *anomaly hunting*;

> "*Always let the science drive the statistics. If you get a statistically significant result, go back and describe what it means in the scientific context.*'"
>
> – Aaron Rendahl

Want more?

Some shameless self-promotion

If you liked this course, check the online materials available at:

`https://github.com/fcampelo/Design-and-Analysis-of-Experiments`

It's free, and I think you'll like it too!

Questions?

Images: http://kawiku.deviantart.com/art/Cousin-Itt-334962933

https://commons.wikimedia.org/wiki/File:Question-mark-blackandwhite.png

About this material

Conditions of use and referencing

Please reference this work as:

Felipe Campelo, *Design And Analysis of Experiments*. Online: http://git.io/vnAnS, International Meeting on Speech Sciences - EICEFALA, September 29, 2015, Belo Horizonte, Brazil; Creative Commons BY-NC-SA 4.0.

```
@Misc{Campelo2015-EICEFALA,
  title={Design And Analysis of Experiments},
  author={Felipe Campelo},
  howPublished={http://git.io/vnAnS},
  year={2015},
  month={September 29},
  note={International Meeting on Speech Sciences - EICEFALA,
        Belo Horizonte, Brazil; Creative Commons BY-NC-SA 4.0.}}
```