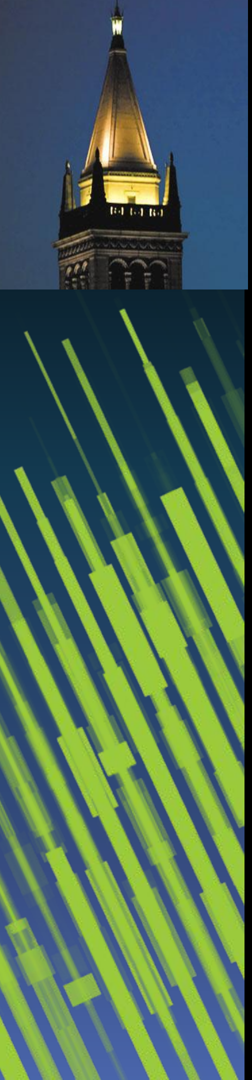


DS 102: Data, Inference, and Decisions

Lecture 15 – Introduction to Design of Experiments

Fernando Pérez
University of California, Berkeley

With thanks to:
Elizabeth Purdom, UC Berkeley
Chris Mack, UT Austin

What is an Experiment?

(noun)

An operation or procedure carried out under controlled conditions in order to discover an unknown effect or law, to test or establish a hypothesis, or to illustrate a known law.

Examples of Experiments?

- Growing cells in media with different concentrations of a chemical (drug).
- Running a numerical model with different parameter choices.
- Changes made to a website (A/B testing).
- A survey? When can surveys be experiments?
- Others?

Causality

- Remember discussion from last two lectures?
- Experiments are our best tool for finding causal relationships.
- We want to *control* our inputs so we can infer what changes in inputs are causing changes in outputs

Experiment maps inputs to outputs

Controlled Inputs (x)

Outputs (y)

Uncontrolled, *observed* inputs (u)

Nuisance Inputs: we'd like to ignore them, but they impact y

Uncontrolled and *unobserved* inputs (v)

Dealing with input variability

- Controlled inputs (x)
  - Systematic variation
- Uncontrolled, observed (u)
  - Blocking: group experiments across reasonably constant values of u
  - Model the impact of u and remove its effect from the model $y=f(x) - g(u)$.
- Uncontrolled, unobserved (v)
  - Randomization: "control what you can, randomize the rest."

Control? What does that mean?

- x are experimentally controlled
- u are partly experimentally and partly statistically controlled
- v are statistically controlled

Repetition and Replication. ???

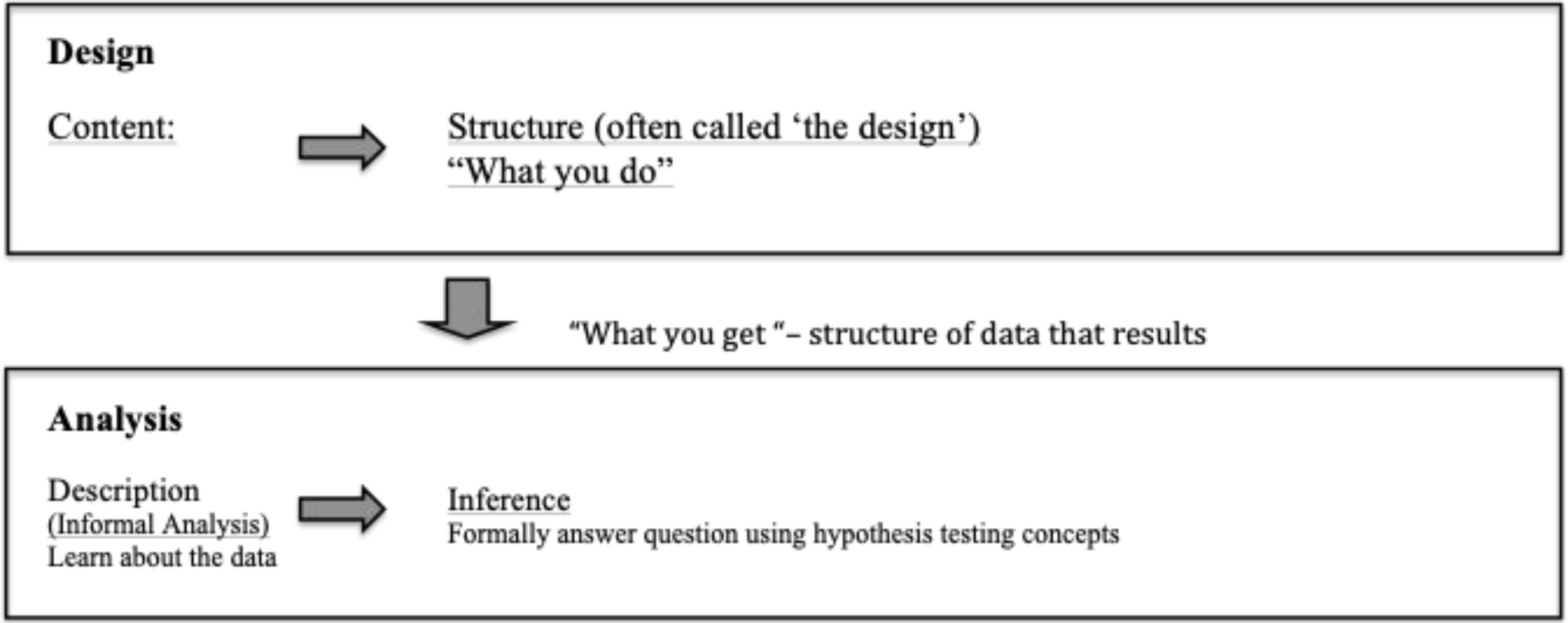
- Replicates: repeated experimental runs, with whole experiment fully repeated.
  - Each replicate independently subject to full variability (say a complete block).
- Repeats: duplicate the experiment on some data within one run.
  - Repeats typically don't re-generate all sources of variation.

Why design experiments?

- Data is expensive!
- Get the most information, knowledge out of every data point.
- Plan the acquisition of data to provide valid conclusions
    - Acquire the data in a way that can lead to good statistical analysis.

An experiment (statistically)

1. Design: choices you make before collecting the data
    a. Driven by a question/hypothesis.
2. Running the experiment: get the data.
3. Analysis: you analyze the data based on your experimental design.


Design
Content:
Structure (often called ‘the design’)
“What you do”
“What you get ”– structure of data that results
Analysis
Description
(Informal Analysis)
Learn about the data
Inference
Formally answer question using hypothesis testing concepts

Content of Experimental Design

- What measurement to make (the response)?
- What conditions to compare (the treatment)?
- What material to apply the treatments to (the units)?

Uses of DoE

- Exploratory work
  - Comparison between alternatives.
  - Screening which factors affect a response.
- Optimize parts of a process
  - Obtain and maintain a target response with minimum variability (control)
  - Max/minimize a target response (output optimization)
  - Reduce overall variability of response (process robustness)
- Regression (modeling)

Blocking

- X, Y: results from Treatment 1, 2
- Measure Z=X-Y?

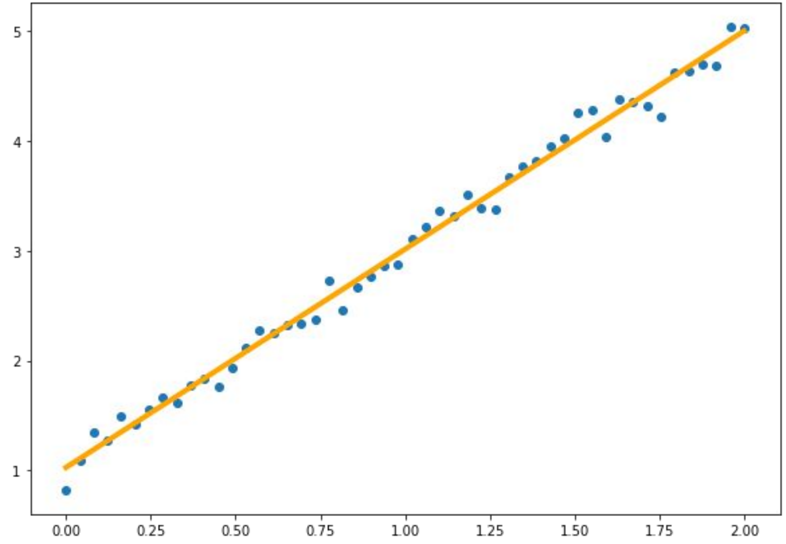
$$\text{var}(Z) = \text{var}(X) + \text{var}(Y) - 2\text{cov}(X,Y)$$

- We can reduce var(Z) by increasing cov(X,Y)
  - An error that is the same for X and Y will cancel!
  - Blocking deliberately increases this covariance.

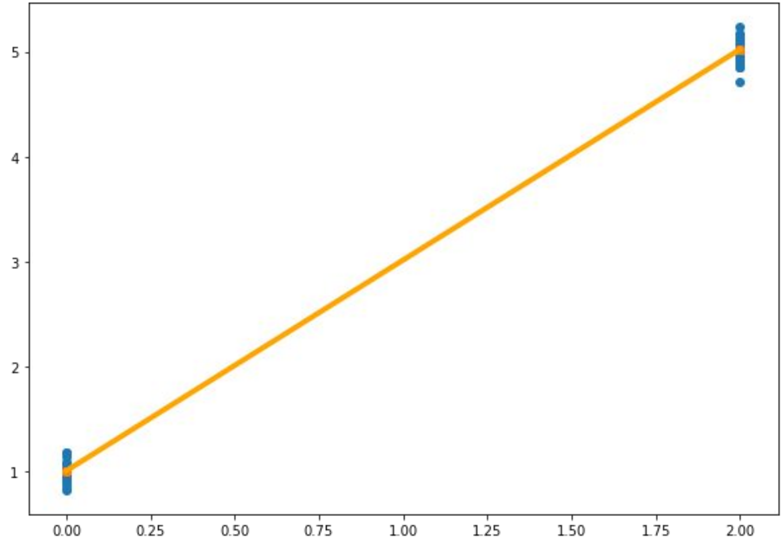
Randomization

- Uncontrolled, unobserved inputs???
  - Randomize to (statistically) control
  - (try to) turn systematic errors into random ones, that (hopefully) average out to zero.
- Note: when dealing with known variation, block if at all possible.

DoE for simple regression



	coef	std err
const	1.0271	0.031
x1	1.9856	0.027



	coef	std err
const	1.0044	0.022
x1	2.0084	0.016

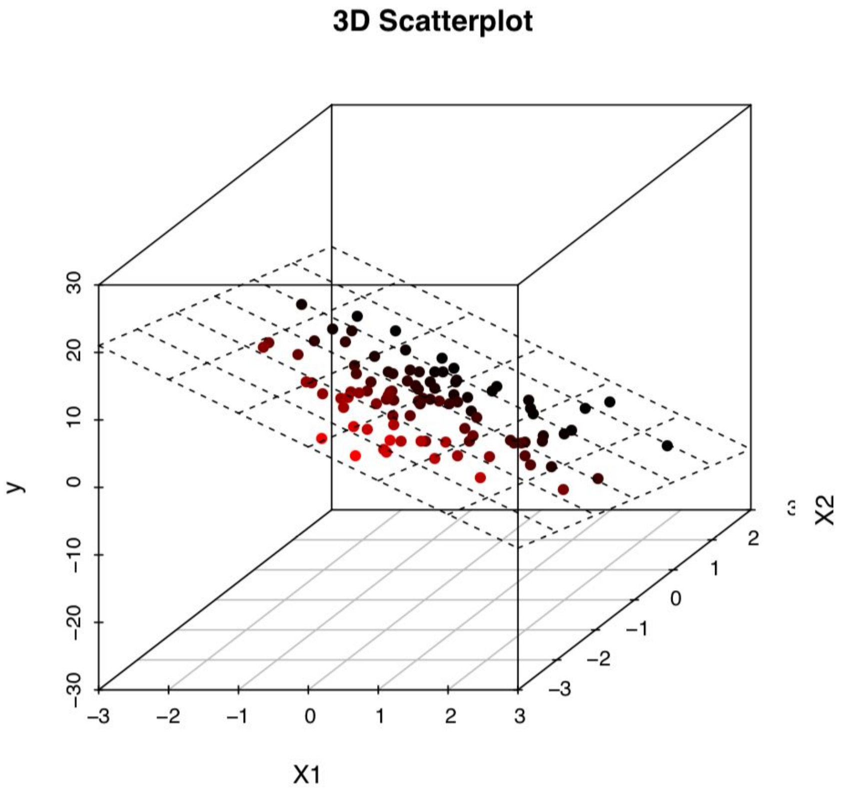
Primary vs Alternate models

- For exploratory work, we may not have a clear idea of what our model could be
- In some cases, we have a clear primary and alternate model in mind
- Simple case: one predictor variable, linear vs. quadratic models
  - Optimizing the design for linear (dumbbell design) means we are insensitive to quadratic variation
  - Optimizing for quadratic gives us reasonable efficiency for a linear model

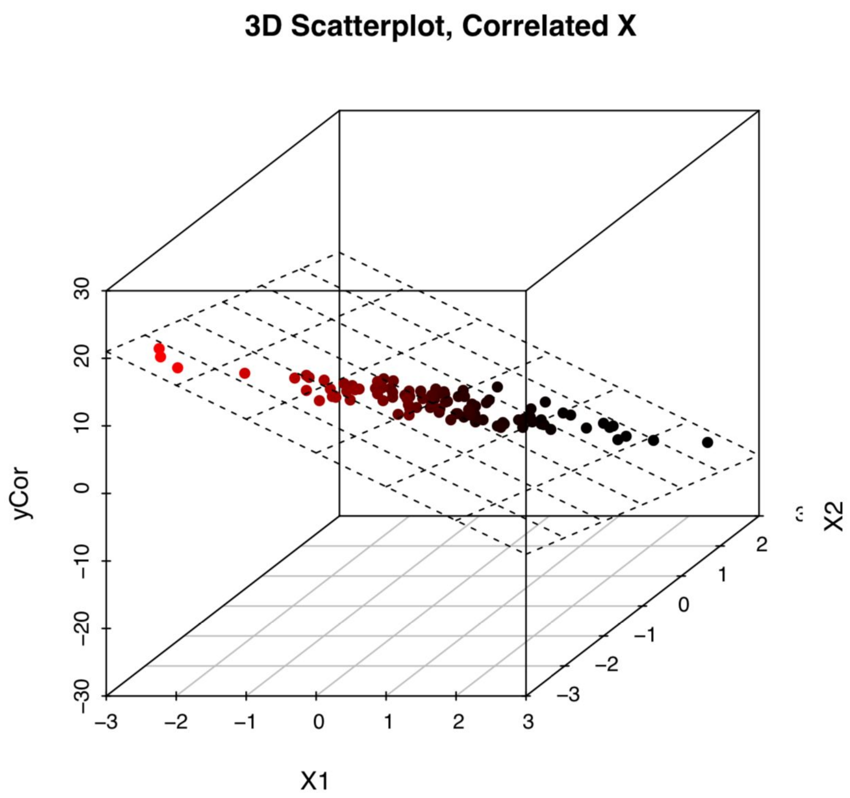
Optimal Design

- General regression model designs can be complicated!
- Optimal Design: algorithmically search design space and optimize a specific statistical metric
  - Non-optimal designs more data to estimate parameters with the same precision
  - Multiple predictor variable? Trade-offs between parameter variances
  - Limitation: model and variable range must be pre-specified. Rigid!

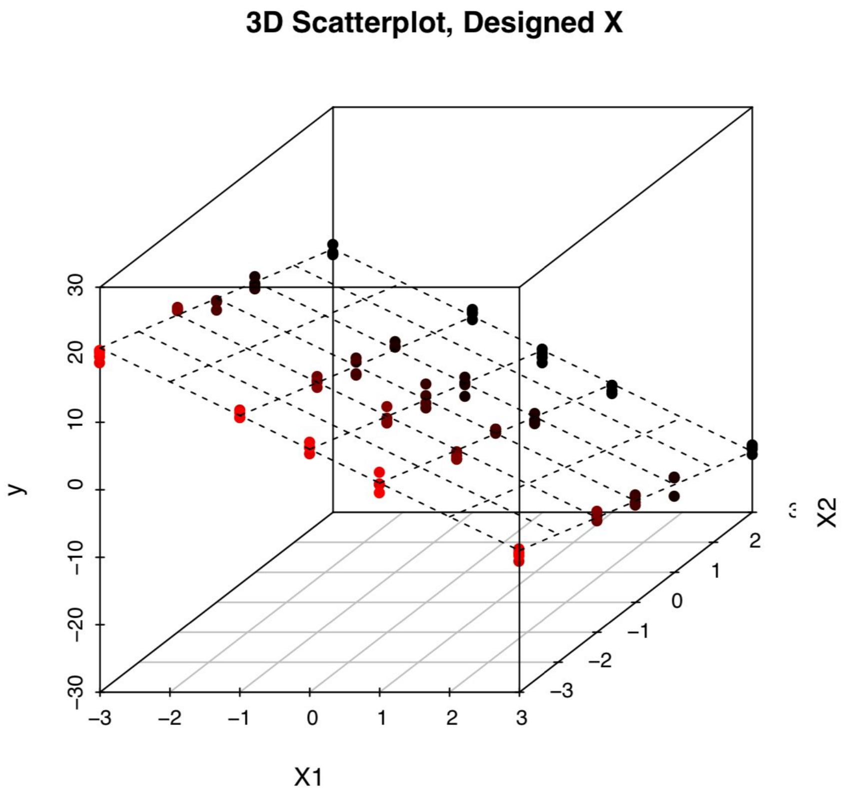
Sampling for regression



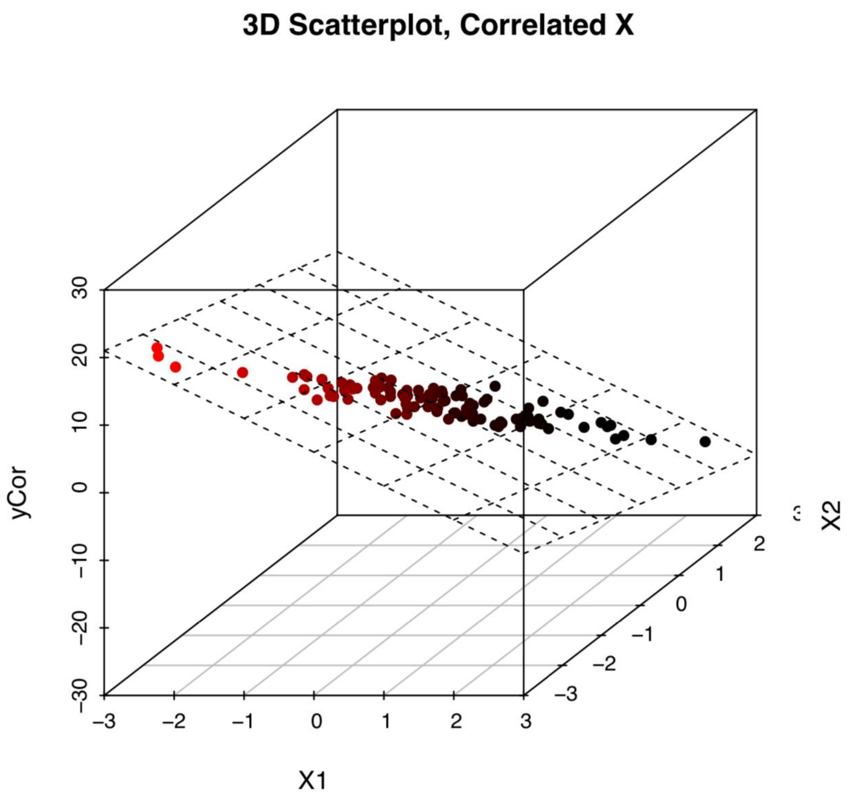
Sampling for regression

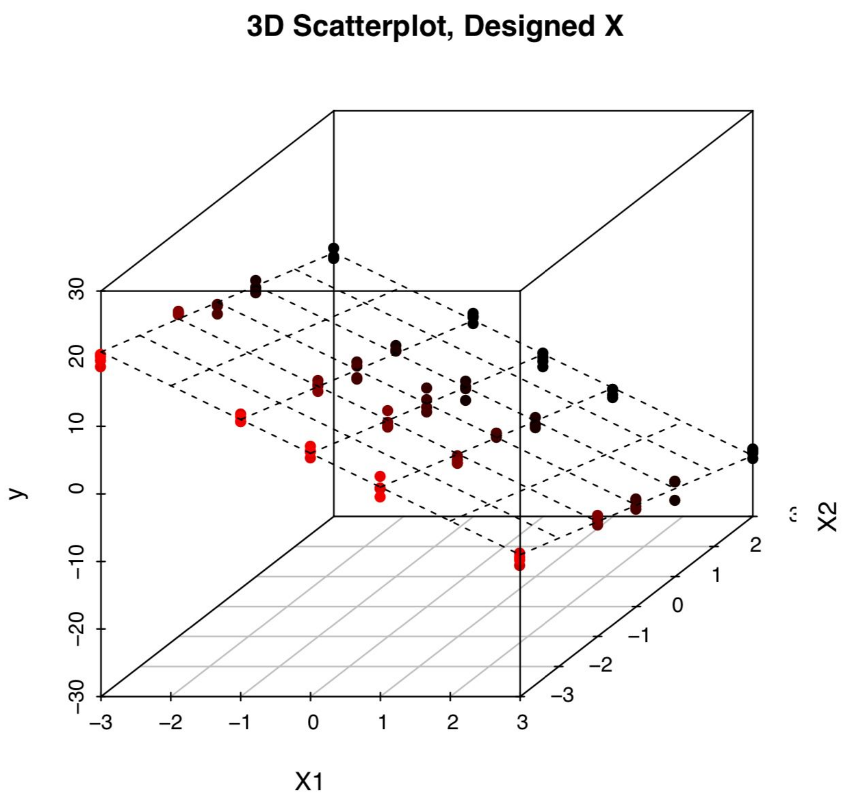


Sampling for regression



Sampling for regression





Advantages of designed experiments

- We pick our design points (and their replication)
  - Avoid collinearity.
- We randomly assign treatments
  - Causality

- Are these two things the same?
- Can't I choose my observational data to be well sampled (esp. if I have a lot)?

Six Principles for Regression Design

(NIST/SEMATECH e-Handbook of Statistical Methods, section 4.3.3)

- Capacity for the primary model
- Capacity for the alternate model
- Minimum variance of estimated coefficients or predicted values
  - Except for simple cases, must search for optimal design
- Sample where the variation is
- Repeats and replication
  - Estimate process standard deviation independent of model
- Randomization and blocking

Analysis follows DoE

- Did you use a block design but not analyze with blocks?
- Better than not having done the blocking!
  - But you're leaving opportunity on the cutting floor.
- Understand your data provenance (D100!)
- Understand how your data was collected
  - Even if you weren't part of the DoE