Tim Book

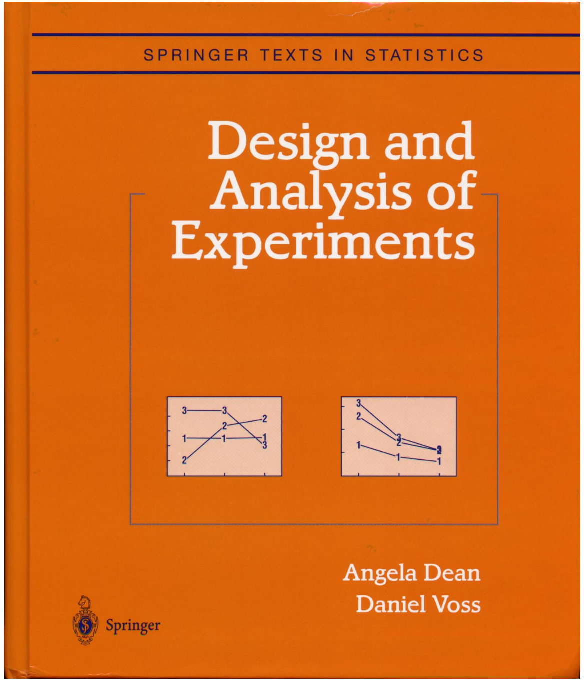
Experimental Design

GENERAL ASSEMBLY

Credit where credit is due

Angela Dean
Emeritus Professor, OSU

Dan Voss
Professor, Wright State



Why would you conduct an experiment?

According to Dean & Voss, there are four reasons:

1. To determine the principal causes in a measured response.
2. To find the conditions that give rise to a maximum or minimum response.
3. To compare the response achieved at different settings of controllable variables.
4. To obtain a mathematical model in order to predict future responses.

Why would you conduct an experiment?

According to Dean & Voss, there are four reasons:

1. *To determine the principal causes in a measured response.*
2. To find the conditions that give rise to a maximum or minimum response.
3. *To compare the response achieved at different settings of controllable variables.*
4. To obtain a mathematical model in order to predict future responses.

ie, "So we can do machine learning on the data."

Why would you conduct an experiment?

According to Dean & Voss, there are four reasons:

1. To determine the principal ***causes*** in a measured response.
2. *To find the conditions that give rise to a maximum or minimum response.*
3. To compare the response achieved at different settings of controllable variables.
4. *To obtain a mathematical model in order to predict future responses.*

We don't actually know how to do this at all yet!

Experimental Design is...

The study of how to conduct the proper experiment given a research question.

"Garbage in, garbage out."

A poorly conducted experiment can lead to an incredible waste of time and money. You may leave your research question unanswerable with the resulting data!

They hate us cuz they ain’t us

Experimental design gives statisticians lots of opportunities to tell natural scientists “no” when conducting experiments. That’s why

statisticians are the scientists scientists love to hate.

How to Design an Experiment

1. Define the objectives of the experiment
2. Identify all sources of variation
3. Choosing a rule for assigning units to treatments
4. Decide on the measurement to be made, experimental procedure, and how you plan to analyze results

“Give me six hours to chop down a tree, and I will spend the first four hours sharpening the axe.”

Abe Lincoln (apocryphal)

Step 1: Define the objectives

You should make a precise list of the questions you'll want to be able to answer after the experiment. Feel free to use the word **cause** here!

Only list the essential questions you'd like to answer.

Too many questions, and you unnecessarily complicate the experiment and increase chances of mistakes.

Too few questions, and the ones you omitted me be unanswerable from your experimental design.

Step 2: Identify all sources of variation

Let's divide all sources of variation into three categories:

- The experimental units
- The ones you care about (treatment factors)
- The ones you don't care about (nuisance variables, blocking factors)

Experimental Units

The **experimental unit** of an experiment is the basic “thing” you’re experimenting on.

If you’re having trouble identifying your experimental unit, try to picture what you want your data to look like at the end of your experiment. Your experimental unit is *probably* what each row represents.

Treatment factors

A **treatment factor** (or just **treatment** for short) is a source of variation that you wish to manipulate manually to measure how it affects the response. A treatment may have two or more **levels**.

In many experiments, you might wish to have a **control treatment level** (the experimental unit receives nothing) and/or a **placebo treatment level** (the experimental unit receives a fake treatment).

Example: The 2008 Obama Campaign

The campaign tested **six different images** and **four different buttons** to see which combination resulted in the greatest number of email addresses given.

There were **two treatment factors** with 6 and 4 **levels**, respectively.

There were thus 24 total treatment combinations

(Actually, there were 26. They tested two other things.)

Example: The 2008 Obama Campaign

The experiment was extremely successful. By using the best combination of image and button, they received 40% more email addresses, from 7.1 million to 10 million.

The average donor gave $21. That's a grand total increase of **$60 million!**

Read more about this here!

The Result?


OBAMA'08
CHANGE
WE CAN BELIEVE IN
JOIN THE MOVEMENT
Email Address
Zip Code
LEARN MORE
PAID FOR BY OBAMA FOR AMERICA
CONTINUE to WEBSITE

Blocks

A **blocking factor** (or **block**, for short) is a source of variation that affects your response in a way that you might consider it a **nuisance variable**. It's not a variable that you can manipulate as part of your experiment, yet it may be interesting to measure the differences between blocks.

Treatments vs. Blocks

Treatment factors are things you administer as part of the experiment. You may care about how treatments affect the response in different **blocks**.

For example, will Medicine A or Medicine B improve a patient's health more? This will differ by patient age, so you separate patients into three categories: "Below 18", "18-54", and "55 and up".

The medicine is the treatment factor, while the age category is a blocking factor.

Example: Cake

(Adapted from Bulletin in Applied Statistics, 1980 by Lewis & Dean as portrayed in Dean & Voss)

The experiments wanted to know how “cake quality” (a 1-10 scale) was affected by adding different amounts of **glycerine** and **tartaric** acid to the cake mix. Measurements were done with three different temperature settings

What is the **experimental unit**?

Example: Cake

(Adapted from Bulletin in Applied Statistics, 1980 by Lewis & Dean as portrayed in Dean & Voss)

The experiments wanted to know how “cake quality” (a 1-10 scale) was affected by adding different amounts of **glycerine** and **tartaric** acid to the cake mix. Measurements were done with three different temperature settings

What is the **response variable**?

Example: Cake

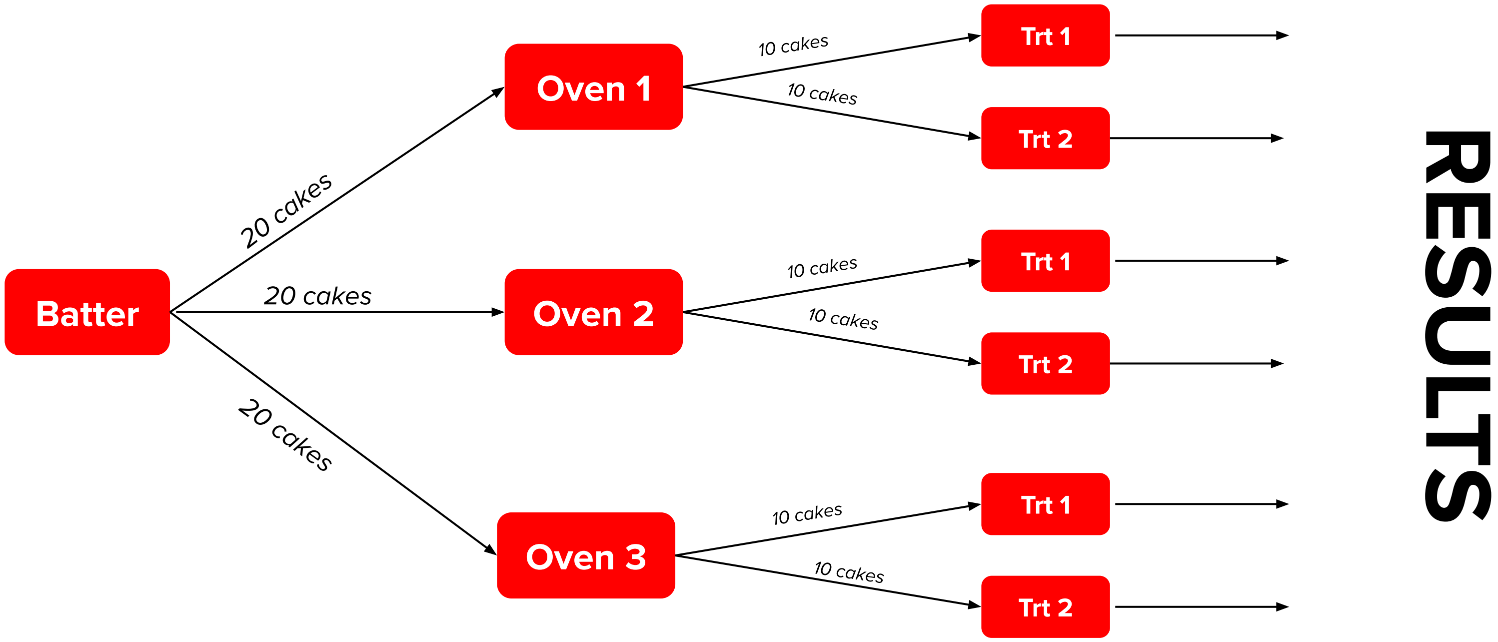
(Adapted from Bulletin in Applied Statistics, 1980 by Lewis & Dean as portrayed in Dean & Voss)

The experiments wanted to know how “cake quality” (a 1-10 scale) was affected by adding different amounts of **glycerine** and **tartaric** acid to the cake mix. Measurements were done with three different temperature settings

What are the **treatment(s)**?

What are the **block(s)**?

Experiment Diagram



Fixed and Random Effects

Sources of variation can fall into two categories:

- **Fixed effects** are variables that the experimenter has specifically selected the factor levels (or there are a limited amount of them) and the experimenter may be interested in finding how the different levels affect the response.
- **Random effects** are variables that have an extremely large number of levels, and the ones surveyed might be a sample of all possible levels. Because of this, we are not interested in comparing how the different levels affect the response.

Huh? An example will help.

You're trying to see which of 3 lightbulb designs lead to longer lasting lightbulbs. The factory producing these lightbulbs has 6 machines that can produce each type, and has to be operated by a technician. The factory has 18 technicians on staff.

What are the **fixed effect(s)**?

What are the **random effect(s)**?

Fixed and Random Effects

A model which has only fixed effects is said to be a **fixed-effects model.**

A model which has only random effects is said to be a **random-effects model.**

A model which has both fixed and random effects is said to be a **mixed-effects model.**

We will deal only with fixed-effects models in this class. Random effects are much more difficult to analyze and conceptualize. Luckily, ignoring potential random effects often does not impact the resulting analyses.

Step 3: Choosing a Rule for Assigning Units to Treatments

When it comes to assigning units to treatments, there is really only one thing to keep in mind:

Step 3: Choosing a Rule for Assigning Units to Treatments

When it comes to assigning units to treatments, there is really only one thing to keep in mind:

RANDOMNESS!!

Step 3: Choosing a Rule for Assigning Units to Treatments

An experiment in which units are randomly assigned to treatments is called a **Completely Randomized Design (CRD)**. If there are blocks involved, you could also call it a **Completely Randomized Block Design (CRBD)**.

Step 3: Choosing a Rule for Assigning Units to Treatments

Randomness is key here since it's the ***ONLY*** way to ensure no unaccounted for factors are affecting your design.

For example: What if a promising new treatment is being tested, but the experimenter subconsciously favors their preferred patients? Or what if they consciously give the new treatment to the patients who are most sick?

If done correctly, CRDs are the ***only way to show causation*** between your factors and your response.

Some things can't be randomly assigned

Sometimes it is impossible or unethical to randomly assign something. For example, if you want to see the effects of smoking on lung health, it would be unethical to randomly assign individuals to various smoking levels.

Similarly, you can never show causation between race and anything, as race cannot be randomly assigned.

Some things can't be randomly assigned

If your treatment (eg smoking) can't be randomly assigned, you can only look at historical data. This is no longer an experiment, but an **observational study**. Observational studies are fine and useful, but they cannot show causation.

Other things, such as sex, cannot be assigned at all. If you're not interested in such a variable as a treatment, **it is probably a block**. In this case, you're still good to go with an experiment.

jupyter To the notebook!

Step 4: Decide on measurement to be made, and plan on how you'll analyze results

How will measurements be made? Take into consideration your sources of variation, and decide on an experimental design. How many treatment levels are there? Are there any blocks? What pitfalls can you run into?

Step 4: Decide on measurement to be made, and plan on how you'll analyze results

You might also consider carrying out **repeated measurements** on an experimental unit if

- If there's potential noise in the measurement itself, it may be wise to take multiple measurements and consider their average.
- Value of response might change with time (in which case, time is now a factor to consider).

Step 4 Side Note

An experiment in which the experimental units are humans, you might be conducting a **survey**. Gathering data from humans is… difficult.

There is an entire subfield of experimental design called **survey methodology** that deals with more than just randomization and blocking factors. It involves reducing survey **bias**, survey **sampling**, and even things like how to properly word survey questions.

Step 4: Decide on measurement to be made, and plan on how you'll analyze results

How will we analyze the results? That'll be the topic of tomorrow's lesson on **A/B Testing** (spoiler alert, this is just another term for **hypothesis testing**).

Before that, though, all analysis should begin with EDA. Let's take a look at some data now.

jupyter To the notebook!

Summary

1. Define the objectives of the experiment
2. Identify all sources of variation
3. Choosing a rule for assigning units to treatments
4. Decide on the measurement to be made, experimental procedure, and how you plan to analyze results

Summary

1. *Define the objectives of the experiment*
2. *Identify all sources of variation*
3. *Choosing a rule for assigning units to treatments*
4. *Decide on the measurement to be made, experimental procedure, and* how you plan to analyze results

Tomorrow's A/B Testing lesson!

