

# Statistics 514: Design of Experiments

## Topic 1

### Topic Overview

This topic will cover

- Fundamentals of Statistics
- Fundamentals of Experimentation
- Alternatives to Experiments
- Context of Experimentation

### What is Statistics?

*Statistics* is the science of problem-solving in the presence of variability.

- “Variability” can also be thought of as “lack of information”
- collection of methods for merging “models” and “data”
- hardly ever one unique solution
  - Different solutions have different strengths/weaknesses.
  - Usually solutions resemble each other.
- report/quantify *evidence*, not *proof*
- basic assumption: can get information on future performance based on past experience
  - Further assumptions are always necessary (often inducing symmetries in the model).

### How to Make Statistics into a Bestseller

*“Knowing what to measure and how to measure it makes a complicated world much less so. If you learn how to look at data in the right way, you can explain riddles that otherwise might have seemed impossible. Because there is nothing like the sheer power of numbers to scrub away layers of confusion and contradiction.”*

Steven Levitt, *Freakonomics*

- *Statisticians* detect patterns and themes in data and models from different fields, in order to develop methodology **that is not specific to any particular field** for understanding data.

“Most scientists study some aspect of nature: rocks, stars, particles; statisticians study scientists, or at least scientific data. Statistics is an information science – the first and most fully developed information science.”

Brad Efron

- *Experimental Design* seeks to realize this promise of generalizability more fully (“statistical truth”).

## What is an Experiment?

**Definition:** An *experiment* is the purposeful setting of an input variable, one or more times, with observation of corresponding output variables.

## Notation

$X$  = input/independent variable (or vector thereof)

$Y$  = output/dependent variable (or vector thereof)

$Y$  is affected to some degree by  $X$ . Extreme cases:

- No relationship (independence)
- Deterministic relationship  $Y = f(X)$ , for some  $f$

We choose values of  $X$ , a run sheet with rows

$$X_1, X_2, \dots, X_n,$$

set  $X$  to these values in turn, and then observe the corresponding values of  $Y$ , getting pairs

$$(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n).$$

(Note that time/space/order is an implicit variable here.)

## Example

5 people get pacemaker and heart drugs ( $X = 0$ )

5 people get heart drugs only ( $X = 1$ )

Measure  $Y$  = regularity of heart beat one month later.

# What is the Purpose of an Experiment?

A *comparative* experiment

- answers questions (tests hypotheses/estimate parameters),
- used to understand and/or improve a system.
- make predictions.

## Goals in Experiments

1. Learn how (or that)  $X$  affects  $Y$ .
2. Find  $X$  with “best”  $Y$ .
3. Solve 1) and/or 2) efficiently.
4. **Manipulation:** “If I set (or see)  $X$  at a certain value, how does that action affect my knowledge of the *distribution* of  $Y$ ?”

“Actions called for as a result of an experiment:

- ‘Cashing in’ on new knowledge.
- Using the new knowledge to look for further possibilities of improvement.”

## Examples

- Improve yield of crops
- Increase wilt resistance of tomatoes
- Make stronger steel
- Reduce variability in size of car door frames

## Advantages of Comparative Experiments

- Experiments allow us to set up a *direct comparison* between treatments of interest.
- We can design experiments to *minimize any bias* in the comparison.
- We can design experiments so that the *error in the comparison* is small.
- **We are in control** of experiments, and having that control allows us to make *stronger inferences* about the nature of differences that we see in the experimental observations. Specifically, we may make inferences about *causation*.

## Scientific Method

Experimentation falls in the larger context of the so-called *Scientific Method*. Provides a context for defining what is “scientific”.

- Observing a phenomenon that is in some way interesting or puzzling. (inspiration)
- Making a guess as to the explanation of the phenomenon. (hypothesis)
- Devising a test to show how likely this explanation is to be true. (experiment)
- Carrying out the test, and on the basis of the results, deciding whether the explanation is a good one or not. (analysis)

## Alternatives to Experiments

- Judgment based on experience
- Rule books
- Mathematical formulas
- Historical data
- Hunches, intuition
- Non-comparative experiments (all the  $X_i$  are the same)
- Investigating accidents

## Judgment Based on Experience

“Based on experience with similar products, we should set the oxidation temperature to 980 degrees.”

- Accuracy depends on person making judgment,
- and how similar this problem really is.
- Memory plays tricks and is selective.

## Rule Books

“To span the given length of a window, the header beam has to be at least a 2 by 6.”

- Carpenters don’t experiment on the site.
- Presumably the drafters of the construction tables did use experiments.

## Mathematical Formula

- Time for an object to fall
  1. Acceleration =  $-9.8\text{m/sec}^2$
  2. So, velocity =  $-9.8tm/\text{sec}$  at time  $t$
  3. So, position =  $-4.9t^2\text{m}$  at time  $t$
  4. So, to fall distance  $d$  takes time  $\sqrt{d/4.9}$ .
- ...approximately [Text formulas may be idealized.]
- Often verified by data.

## Hunches, Intuition

- Experimental design can't replace ingenuity.
- But it can leverage it.

## Non-comparative Experiments

- Example: medical studies with “historical controls”
- Problem: hidden confounding

## Investigating Accidents

- Post-tragedy: investigate causes
- Accidental discoveries: ivory soap, x-rays, penicillin.

## Observational Studies

- Surely the past  $(X, Y)$  data are informative?
- Yes (sometimes)
- But
  1. Past correlations don't imply causal link (experiments can)
  2. Data may be inconsistent (new machines, new people, new specs)
  3. Historical  $X$ 's may be in too narrow range.

In an **experiment**, you start with one set of subjects or material to which you **assign treatments** (conditions you want to compare).

In an **observational study**, you start with several **populations** of subjects or objects (with the conditions already built in); you *take samples* from populations.

An experiment compares treatments.  
An observational study compares populations.

“It is much easier to isolate the effects of interest if you can assign conditions. In an observational study, the conditions you want to study will almost never be the only thing that makes one population different from another. This makes it hard to identify the effects responsible for observed differences.”

### Study of studies

“Statistics, second edition” by Friedman, Pisani, Purves, and Adhikari has the following example. For randomized experiments compared to ones with historical controls, results reported  $\pm$ .

Therapy	Exp +	Exp -	Hist +	Hist -
Coronary bypass surgery	1	7	16	5
Anticoagulants	1	9	5	1
5-FU (a chemo treatment)	0	5	2	0
BCG (for melanoma)	2	2	4	0
DES (to prevent miscarriage)	0	3	5	0

Also, for coronary bypass surgery, 3-year survival for the “control group” was better in randomized experiments than in historical control experiments. Picking the comparison group is tricky.

### Categories of Experimental Problems

1. *Treatment Comparisons* - Compare several treatments and select the best ones.
2. *Variable Screening* - Which variables are important/have an effect on the response?
3. *Response Surface Exploration* - What is the relationship between the response and the predictors?
4. *System Optimization* - Interested only in the region containing the optimum settings of the variables.
5. *System Robustness* - Choose control factor settings at which the system is less sensitive to noise variation.

# Approach to Planning an Experiment

## 1. Statement of problem/objective

- Should be unbiased, specific, and measurable
- What is the experiment intended to do? Also, does anybody care?
- Use relevant background (theory/expertise/experience) to decide how experiments fit into the study of the process or system.
- Obvious step but often overlooked. Sound questions go a long way toward solution.
- Don't jump to technical problem too quickly – focus should be on narrative (problem/solution) rather than difficult technical questions.

## 2. Choose measurement to make (response)

- Are variables measurable?
- What sort of response is expected?
- How accurately can response be measured?
- What is the operational/measurable range?
- Reliability and validity: how good is the response?
  - **Reliability** and **validity** both refer to the soundness of the connections between the goal of an experiment and your choice of response, that is, between the kinds of conclusions you hope to draw and the kind of evidence you plan to gather.
  - Reliability is concerned with repeatability.  
Validity is concerned with relevance of your data (response) to the purpose of your study (and is often hard to establish).

## 3. Choose conditions to study (treatment factors and levels).

- What inputs may affect response? What inputs are of interest?
- Are factors to be held constant? Varied at specific levels?
- What interactions are known/suspected?

## 4. Choose experimental plan (design stage)

- How large a difference in response is practically important?
- How much variation is present?
- What costs and resources are available? What is the timing of the experiment?
- Is the whole experiment to be randomized? Are different factors randomized differently?

5. Perform the experiment

- Make sure the design can be implemented.
- Sometimes necessary to perform a trial run.
- Record deviations from planned experiment.

6. Analyze the data

- Analysis is not part of design, but consider it during planning.
- Do the model assumptions hold?
- This is where academic statisticians come in.
- *Internal validity* – logic of experiment
  - Does it make sense that the treatment produced the measured effect?
  - Extraneous variables (“threats to validity”) might confound effect
- *External validity* – proposed interpretation of study results
  - How well does this generalize?

7. Draw conclusions and make recommendations.

- Conclusions should refer back to the stated objectives of the experiment.
- A *confirmation* or *follow-up experiment* might be necessary.

## Planning, Conducting, and Analyzing an Experiment (Notes)

- **Get *statistical thinking* involved early.** Randomization provides justification for usual  $F$ -test analysis. Helps avoid unintentional subjective biases in assignments.
- **Your *non-statistical knowledge* is crucial to success.** Statistical techniques are most effective when combined with problem-specific knowledge. Ask questions to discover as much about the problem as possible.
- **Keep the design and analysis simple.** Can often answer questions with sound, straightforward approach. Complex designs more sensitive to problems.
- **Practical vs. statistical significance:** Need to initially consider what is “important” difference. Helps determine appropriate sample size. A statistical difference may not be anything of value.
- **Think and experiment iteratively.** Often start with little knowledge of problem and variability. Pilot studies can be done to obtain information and/or used to ensure experiment can be run as planned. Additional experiments may focus on new levels of important factors or include a new factor.

## Reality Check: Montgomery's Theorems

- **Theorem 1.** If something can go wrong in conducting an experiment, it will.
- **Theorem 2.** The probability of successfully completing an experiment is inversely proportional to the number of runs.
- **Theorem 3.** Never let one person design and conduct an experiment alone, particularly if that person is a subject-matter expert in the field of study.
- **Theorem 4.** All experiments are *designed* experiments; some of them are designed well, and some of them are designed really badly. The badly design ones often tell you nothing.
- **Theorem 5.** About 80 percent of your success in conducting a designed experiment results directly from how well you do the pre-experimental planning.
- **Theorem 6.** It is impossible to overestimate the logistical complexities associated with running an experiment in a “complex” setting, such as a factory or plant.

“Without good experimental design, we often end up doing PARC (Planning After the Research is Complete) Analysis.”