

# Module 1 : History and basic principles of design of experiments

Marie-Abèle Bind

Design of Experiments - STAT 140

Spring 2026

# 1918 to 1950 : The agricultural origins (1/2)

- 1920s-1930s : R. A. Fisher and his colleagues at Rothamsted Experimental Station, one of the oldest agricultural research institutions (founded in 1843).
  - Many scientists have worked at Rothamsted (e.g., Yates and Cochran), which many consider as the most important birthplace of modern statistical theory and practice.
- Profound impact on agricultural science (e.g., increase crop yield and development of 2,4-D weed-killer during WWII).
- Factorial designs : experiment in which each combination of controllable factors is considered at several levels (Fisher and Yates).
  - If one has four factors involving five levels each, a full-factorial experiment would require  $5^4 = 625$  distinct observations.
  - Can be impractical : elegant mathematical theory of incomplete block designs such as fractional factorial (Finney, 1945).

## 1918 to 1950 : The agricultural origins (2/2)

- ANOVA : observed variance in a particular variable is partitioned into components attributable to different sources of variation.
  - Fisher introduced the term variance in 1918.
  - Computationally elegant statistical method.
- Clinical trials : The British Medical Research Council officially recognized their importance from the 1930s.
  - The first randomized curative trial was carried out by Sir G. Marshall (1887-1982).
  - The trial, carried out between 1946-1947, aimed to test the efficacy of the chemical streptomycin for curing pulmonary tuberculosis.
  - The trial was both double-blind and placebo-controlled.

# 1951 to late 1970s : The first industrial era

- Box's and Wilson's work on response surface designs (1951)
  - Simple linear or quadratic functions of independent variables that are used to approximate a more complex relationship between a response and independent variables.
  - Typically, the goal in many industrial experiments is to identify the important factors that affect one or more responses from among a large set of factors.
  - The high cost of industrial experimentation limits the number of runs; hence fractional designs with factors typically at two levels are used in these experiments.
  - Once a smaller set of important factors has been identified, the response surface can be studied more thoroughly using designs with more than two levels, and product performance can be optimized.
- Applications in the chemical and process industries
- 1959 : Chernoff describes a procedure for the sequential design of experiments where the problem is hypothesis testing.

## Late 1970s to 1990 : The second industrial era

- Quality improvement initiatives in many companies : systematic and formal approach to the analysis of practice performance and efforts to improve performance
- G. Taguchi's work on robust design for variability and bias reduction using response surfaces to enhance quality and reliability
  - Variability : identifying important "noise" variables
  - Toyota is an early adopter of his ideas
  - Collaborated with Fisher and Rao (Indian Statistical Institute), Tukey (Bell Labs)
- Many applications
  - Automobile
  - Semiconductor
  - Nuclear

# Beginning 1990 : The modern era

- Economic competitiveness and globalization is driving all sectors of the economy to be more competitive
- Statistical Methods for Testing and Evaluating Defense Systems : Interim Report, 1995
  - "Some basic insights of experimental design have had revolutionary impact, but many of these insights are not well known among scientists without specialized training in statistics, partly because elementary texts and first courses seldom allocate time to this topic at all, or with any depth. For example, the role of randomization and the inefficiency of the practice of varying one factor at a time are not well appreciated. [...] Much of the support for research in experimental design derived from problems faced by DoD during and shortly after World War II. [...] What is required is the kind of expertise that can adapt underlying basic principles to the current situation, an expertise rarely found outside the scope of well-trained statisticians who understand the relation of standard rules to underlying principles."

# Basic principles of experimental designs

# The need for a control

- In comparing two systems, a new one and a standard one whose behavior is relatively well known, there used to be a natural tendency to test and evaluate the new system separately.
- To avoid this bias, it is common place to test both systems simultaneously under similar circumstances.
- The outcome of an experiment depends not only on the overall quality of the system, but also on more or less random variations, some of which are due to the general environment.
- In selecting the settings of the controls in a experiment, an experimenter must use some background information on what to expect.

- If natural variations in the environment have a relatively large effect on the variability in performance, the ability to match pairs has a correspondingly large effect on increasing the precision of conclusions.
- This principle of matching is generalized to more than two systems, it is referred to as blocking, a term derived from agricultural experiments, in which several treatments are applied in each of many blocks of land.
- How should the various treatments be distributed within a block ?
  - In an agricultural experiment, if one assumes that position within the block has no effect, position will not matter. But if there is a systematic gradient in soil fertility in one direction, the use of a systematic allocation might introduce a bias.

# Randomization

- Another approach to reducing the bias is to select the allocation within the block by randomization.
- Often in operational testing applications with a small number of test articles, randomization may not be necessary, and small systematic designs can be used safely. Or one can select a design at random from a restricted class of reasonably safe designs.
- Randomized experimental designs have traditionally been viewed as the most credible basis for causal inference, as reflected in the typical reliance of the U.S. Food and Drug Administration on such experiments in its approval process for pharmaceutical treatments.

# Replication

- Replication can be viewed as the heart of all of statistics.
- Basic issue behind certain methods we will learn in order to get a handle on how precise our estimates are at the end.
- Some perspectives want to estimate or control the uncertainty. Can be achieved through replication.

# Confounding (1/3)

- Confounding is something that is usually considered bad. Here is an example. Let's say we are doing a medical study with drugs A and B. We put 10 subjects on drug A and 10 on drug B. If we categorize our subjects by gender, how should we allocate our drugs to our subjects? Let's make it easy and say that there are 10 male and 10 female subjects. A balanced way of doing this study would be to put five males on drug A and five males on drug B, five females on drug A and five females on drug B. This is a perfectly balanced experiment such that if there is a difference between male and female at least it will equally influence the results from drug A and the results from drug B.

## Confounding (2/3)

- An alternative scenario might occur if patients were randomly assigned treatments as they came in the door. At the end of the study, they might realize that drug A had only been given to the male subjects and drug B was only given to the female subjects. We would call this design totally confounded. This refers to the fact that if you analyze the difference between the average response of the subjects on A and the average response of the subjects on B, this is exactly the same as the average response on males and the average response on females. You would not have any reliable conclusion from this study at all. The difference between the two drugs A and B, might just as well be due to the gender of the subjects since the two factors are totally confounded.

## Confounding (3/3)

- Confounding is something we typically want to avoid but when we are building complex experiments we sometimes can use confounding to our advantage. We will confound things we are not interested in order to have more efficient experiments for the things we are interested in. This will come up in multiple factor experiments. We may be interested in main effects but not interactions so we will confound the interactions in this way in order to reduce the sample size, and thus the cost of the experiment, but still have good information on the main effects.

# Factorial designs

- Varying one factor at a time is inefficient : the resulting estimates have higher variance than estimates derived from experiments with the same number of replications in which several factors are simultaneously varied.
- It is much better to design an experiment that simultaneously includes combinations of multiple factors that may affect the outcome. Then you learn not only about the primary factors of interest but also about these other factors : blocking variables or factors that may help you understand the interactions or the relationships between the factors that influence the response).
- Optimal experimental designs : allow estimation of causal estimand without bias and with minimum variance.