Topic 1 Introduction to the Principles of Experimental Design

Jan 23, 2025

Experiment

An exercise designed to determine the effects of one or more variables (treatments) on one or more characteristics (response variables) of some well-defined system (experimental unit)

Where does analysis enter the experimental process?

The "Scientific Method"

Formulate a hypothesis
Plan an experiment
Conduct the experiment
Analyze and interpret the results

What the statisticians have to say...

"...the time to think about statistical inference...is when the experiment is being planned." - G Cochran and W Cox

"Statisticians make their most valuable contributions if they are consulted in the planning stages of an investigation. Proper experimental design is often more important than sophisticated statistical analysis." - GJ Hahn

"When I'm called in after it's all over, I often feel like a coroner. I can sign the death certificate—but do little more." - H Ginsburg

"You cannot save by analysis what you bungle by design." - R Light et al.

Steps in experimentation

Define the question
Determine the intended scope of conclusions
Identify the appropriate design/analysis

—

Select the experimental material
Select the treatments
Select the sampling unit and number of replications
Ensure proper sampling
Ensure proper means of data collection
Collect the data
Analyze the data
Interpret the results

—

Effectively communicate the process, data, and results

Experimental Design

The logical structure of an experiment

Hypothesis, scope, and experimental design

The selection of an experimental design depends on your **objective** (hypothesis + scope).

Variety screening trial: Maximize accession number (less replication)
Variety release trial: Appropriate replication for high precision

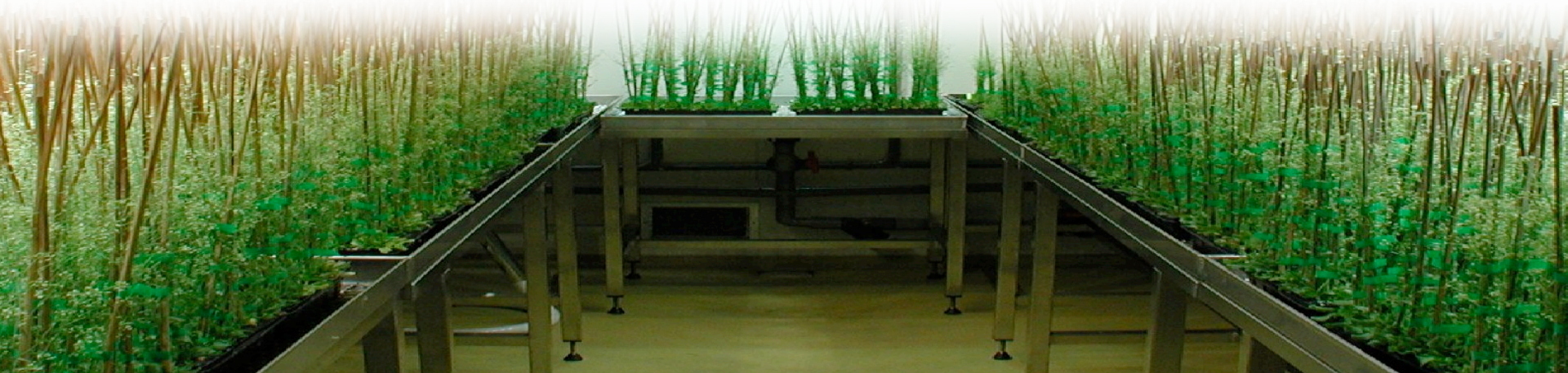
The intended scope of conclusions is a major determinant of experimental design.

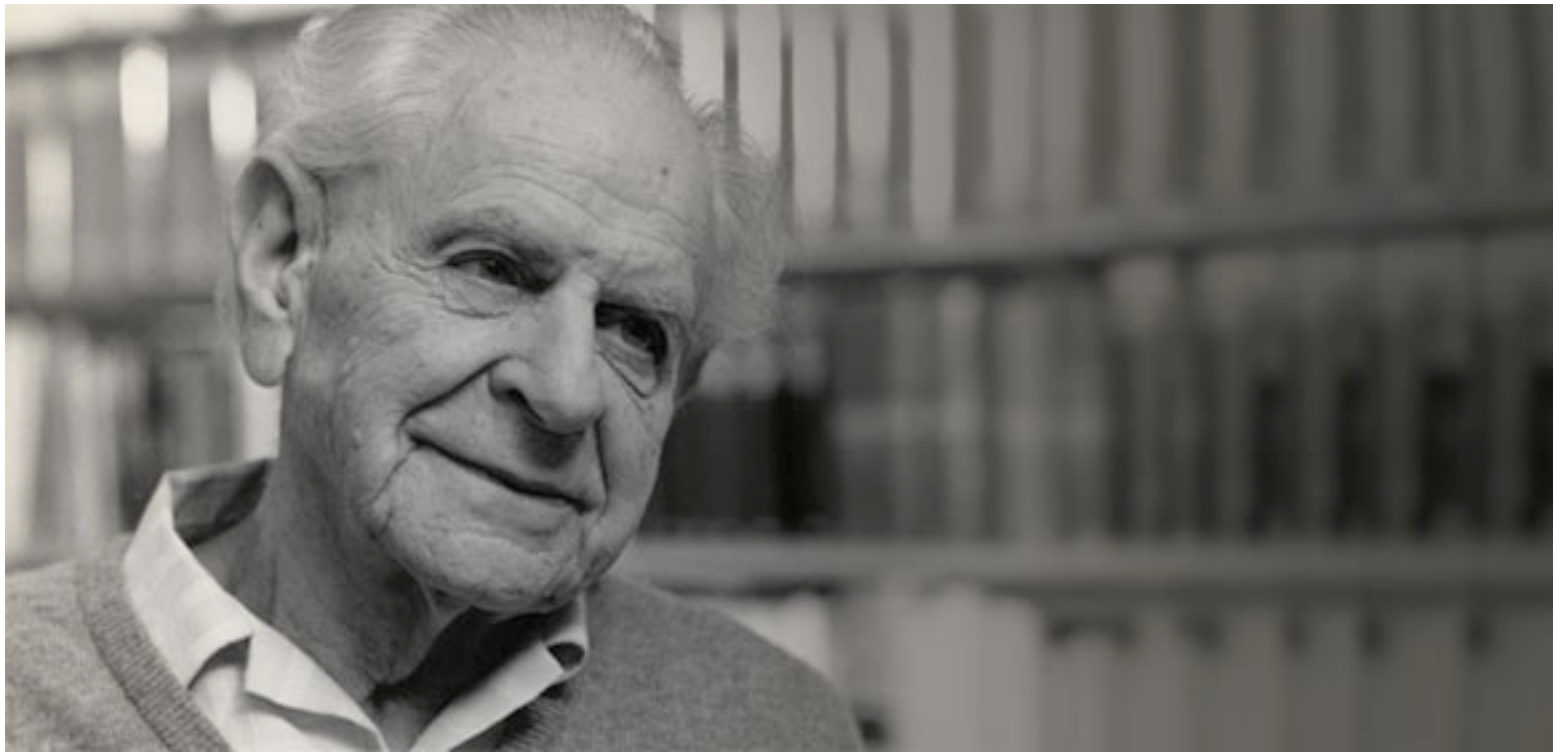
Variety release trial for northern New Hampshire
Variety release trial for New England

Descriptive Science

vs.

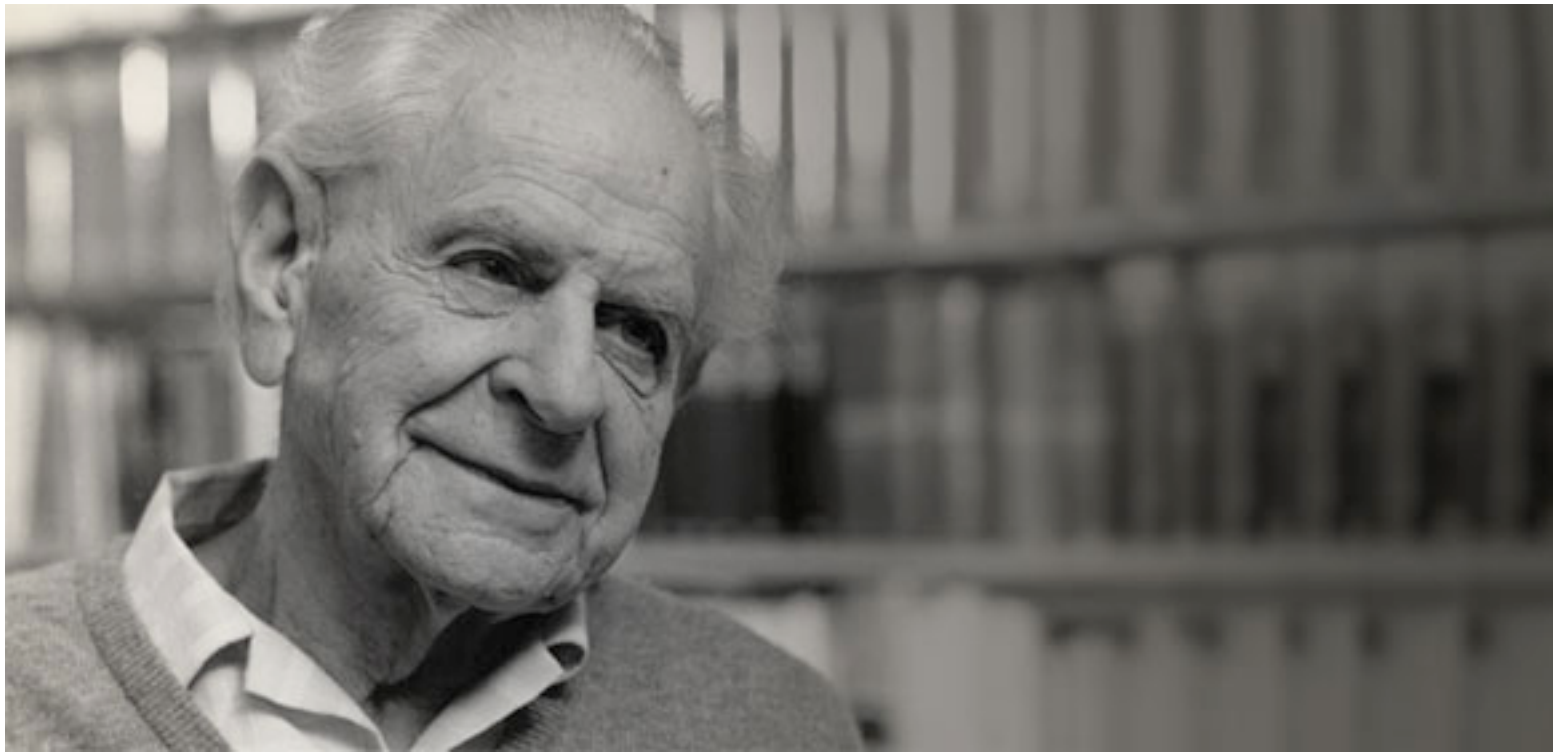
Hypothesis-Driven Science





Conjecture and Refutation

The formulation of informative (risky) ideas
and
the sincere, methodical attempt at their falsification



“Clearly the instruction, ‘Observe!’ is absurd. ...Observation is always selective. It needs a chosen object, a definite task, a point of view, a problem.”

Conjectures and Refutations: The Growth of Scientific Knowledge

Concepts about hypotheses

A hypothesis is a statement that can be tested and falsified.

A null hypothesis (H_0) can never be proven correct. It can only be rejected with known (chosen) risks of being wrong.

A good experimental design allows the **quantification of uncertainty**. The experiment is designed so that it one can calculate the possibility of obtaining the observed results by chance alone.

Testing hypotheses

The significance level (α) is the probability that one rejects H_0 when it is true (Type I error rate). For many fields, default α = 0.05 (1 error in 20).

Lowering α protects you from false positives but increases the Type II error rate (β), the probability of failing to reject H_0 when it is false.

H_0	is rejected	is not rejected
is true	**Type I error**	Correct decision
is false	Correct decision	**Type II error**

The only ways to reduce both types of error simultaneously are to:

1. Increase the number of replications
2. Improve the experimental design

Experimental Design

The logical structure of an experiment

Treatment structure
Replication
Design structure
Response structure
Error control

1. Treatment structure

The set of treatments used and how they relate to one another.

2. Treatment replication

The number of experimental units to be subjected to each treatment.

3. Design structure

How treatments are assigned to experimental units.

4. Response structure

The set of response variables to be measured and the sampling plan that specifies when, where, and with what components of the experimental units one will measure the response variables.

5. Error control

"Noise" reduction through the strategic use of proper protocols, blocking techniques, covariables, or environmental controls (e.g. growth chambers, greenhouses, lab studies).

Among all the previous considerations,
REPLICATION and RANDOMIZATION
are the most important basic principles
in designing experiments

Replication: the number of experimental units that are treated alike.

Experimental unit (e.u.): The smallest system or unit of experimental material to which a single treatment (or treatment combination) is assigned and which is dealt with independently of other such systems under that treatment at all stages in the experiment in which important variation may enter. - Hurlbert 2006

The functions of replication

σ^2_{ε}

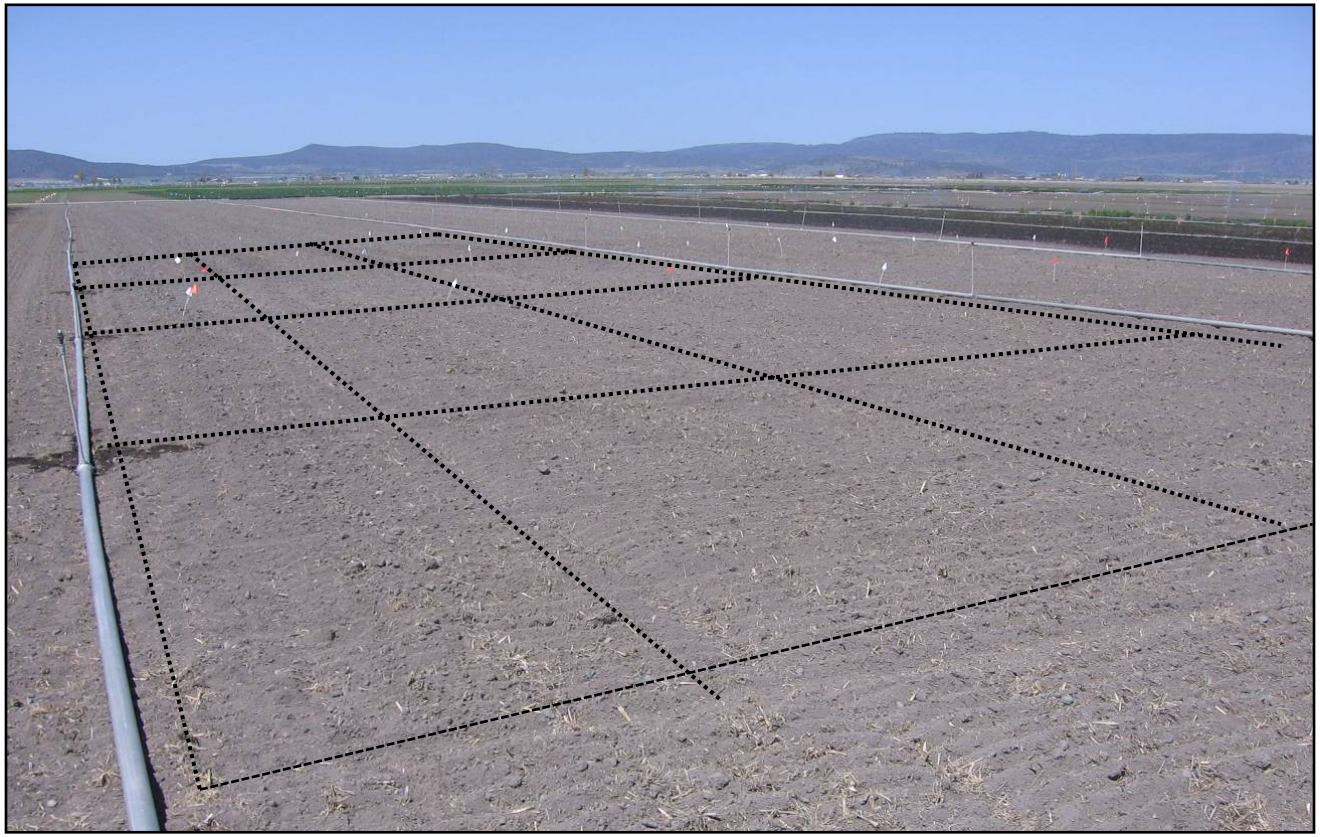
1. To provide an **estimate of experimental error**
2. To improve **precision**
3. To **increase the scope of inference**

$$\sigma_{\bar{Y}(n)} = \frac{\sigma}{\sqrt{n}}$$

Each replication is <u>independent</u> from every other

Example: A field trial comparing three different fertilizers (A, B, C). Response variable: Yield (kg).

1	2	3	4
5	6	7	8
9	10	11	12



Case 1: Four plots are selected at random to receive each fertilizer. The total yield of each plot is measured at the end of the season.

B	A	C	B
C	A	B	A
C	B	A	C

This is a completely randomized design (CRD) with four replications per treatment. If the experiment were repeated the following year with a new random assignments of treatments, this could provide a second set of replications.

Case 2: Same as Case 1, but now the crop is a perennial. The total yield of each plot is measured at the end of each season.

B	A	C	B
C	A	B	A
C	B	A	C

In this case, the plots in the second year are not replications; they are repeated measurements.

Case 3: Same as Case 1, except each of the 12 plots is further divided into three subplots and yield is measured separately for each subplot.

B	B	B	A	A	A	C	C	C	B	B	B
C	C	C	A	A	A	B	B	B	A	A	A
C	C	C	B	B	B	A	A	A	C	C	C

The experimental unit is still the plot. The subplots are not replications, they are subsamples.

Case 4: The three treatment levels are randomly assigned to the three rows in the field. Yield is measured on each plot.

B	B	B	B
A	A	A	A
C	C	C	C

In this case, the experimental unit is the row. Each plot is a subsample. This experiment has no replication.

If you are unsure what the experimental unit is, ask yourself:

"WHAT WAS RANDOMIZED?"

Types and number of measurements

The theoretical consideration

$$\sigma_{\bar{Y}(n)} = \frac{\sigma}{\sqrt{n}}$$

The standard error determines the lengths of confidence intervals and the powers of tests. Decreasing the standard error increases the precision of the experiment.

Precision has to do with the concept of random errors, and the precision of an average can always be improved by increasing the sample size (n).

A good experimental design has **sufficient precision**, meaning there is a high probability that the experiment can detect expected differences.

Types and number of measurements

The practical consideration

Finite resources

Subsampling

Multiple measurements on the same experimental unit can yield a more precise value for that experimental unit. This can reduce the apparent variation among experimental units subjected to the same treatment and thereby reduce the standard error.

$$\sigma_{\bar{Y}(n)} = \frac{\sigma}{\sqrt{n}}$$

A strategic combination of replications and subsamples can be used to achieve the necessary precision, given limited resources.

Other forms of error control

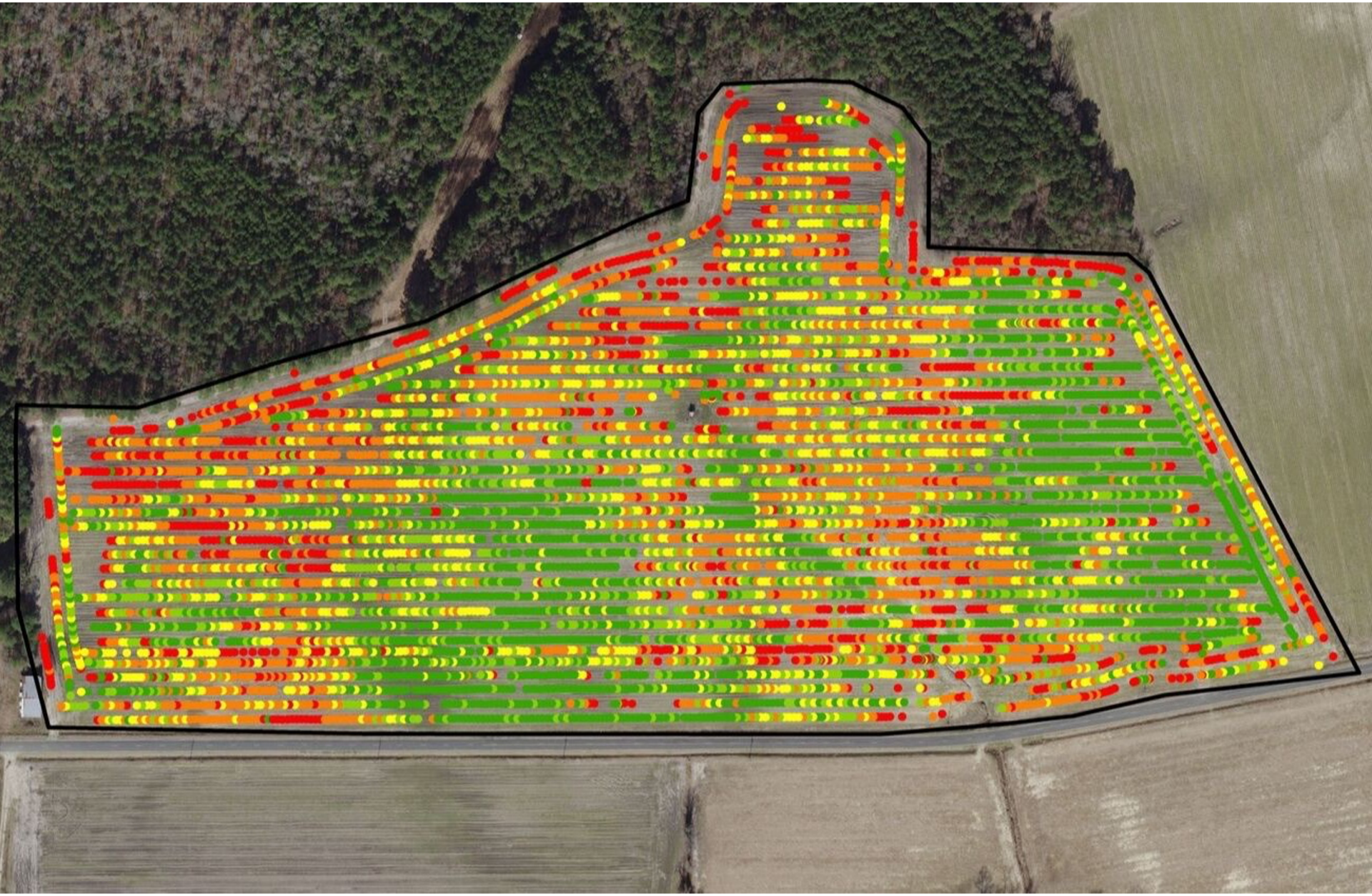
A good experimentalist is always looking for ways to reduce unwanted variation.

SIGNAL >> NOISE

Some available strategies:

1. **Blocks**
2. **Covariables**
3. **Better protocols**

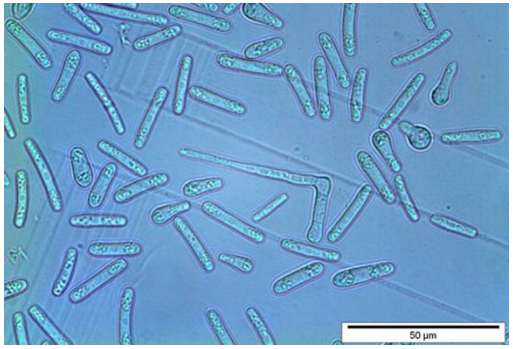
Homogeneous experimental materials
Good equipment
Careful observation
Well-trained people



https://www.antedis.com/fonctions_ajax2.php?lg=en&numphoto=573&nomphoto=micro-parcelles-detourees-experimentation-antedis.jpg&modalPhoto=o&titrePhoto=Cropping&legendePhoto=Cropping+leads+to+clean+micro-plot+delimitation


ARGUS
ARGUS
ARGUS

https://www.pnas.org/content/116/37/18149



$$\sigma^2_\varepsilon$$

Another practical consideration

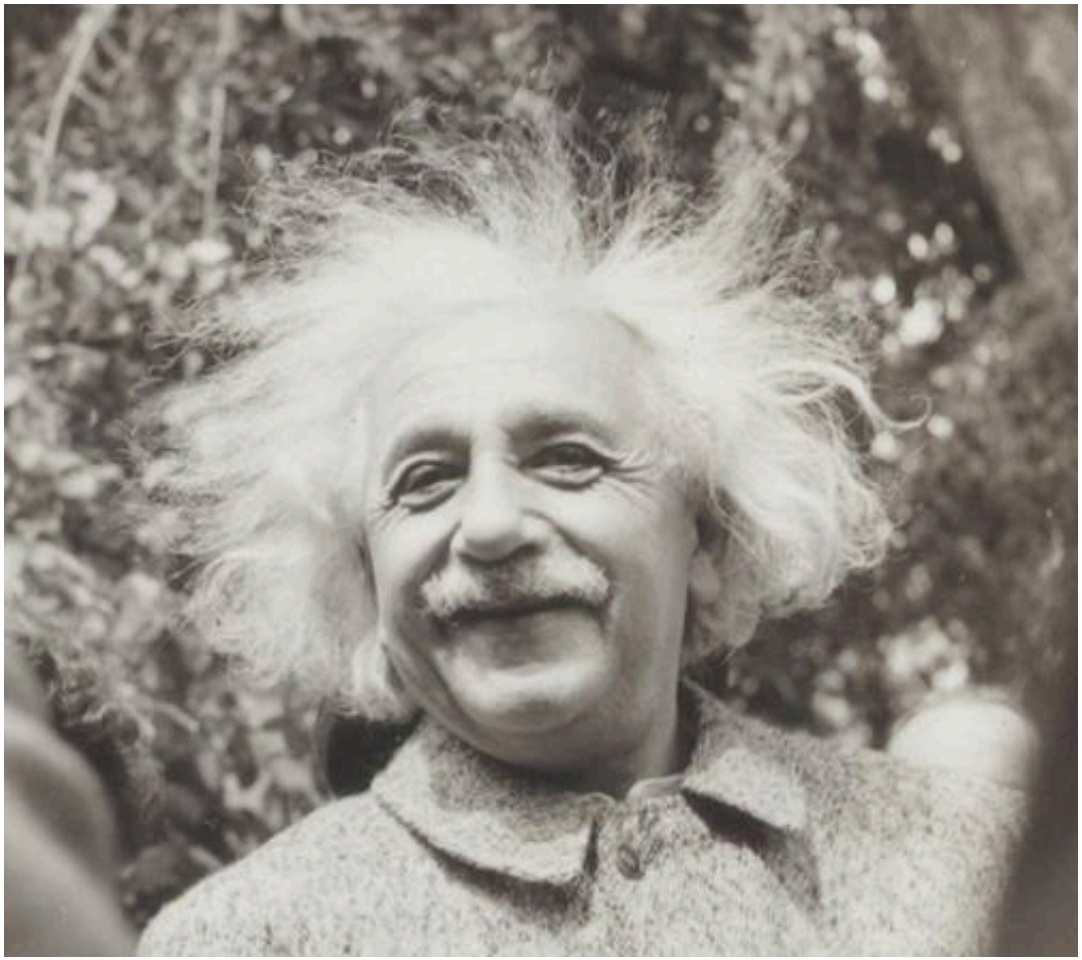
Statistical significance ≠ Biological significance

Too many measurements without real purpose can lead to the statistical declaration of significance, even in the absence of meaningful biological effect:

With a 5% Type I error rate and 200 comparisons, one can expect about 10 false positives (Type I errors). Thus it is important to be clear and strategic about your experimental objective.

With sufficient replication, a difference can always be found:

With intense sampling, one could show that some resource-intensive mitigation effort reduces active N in surface waters by 1 ppm. But who cares? It is important to define "biological significance" and then design experiments to detect that amount, no more and no less.



Not everything that counts can be counted,
and not everything that can be counted counts.

Design structure

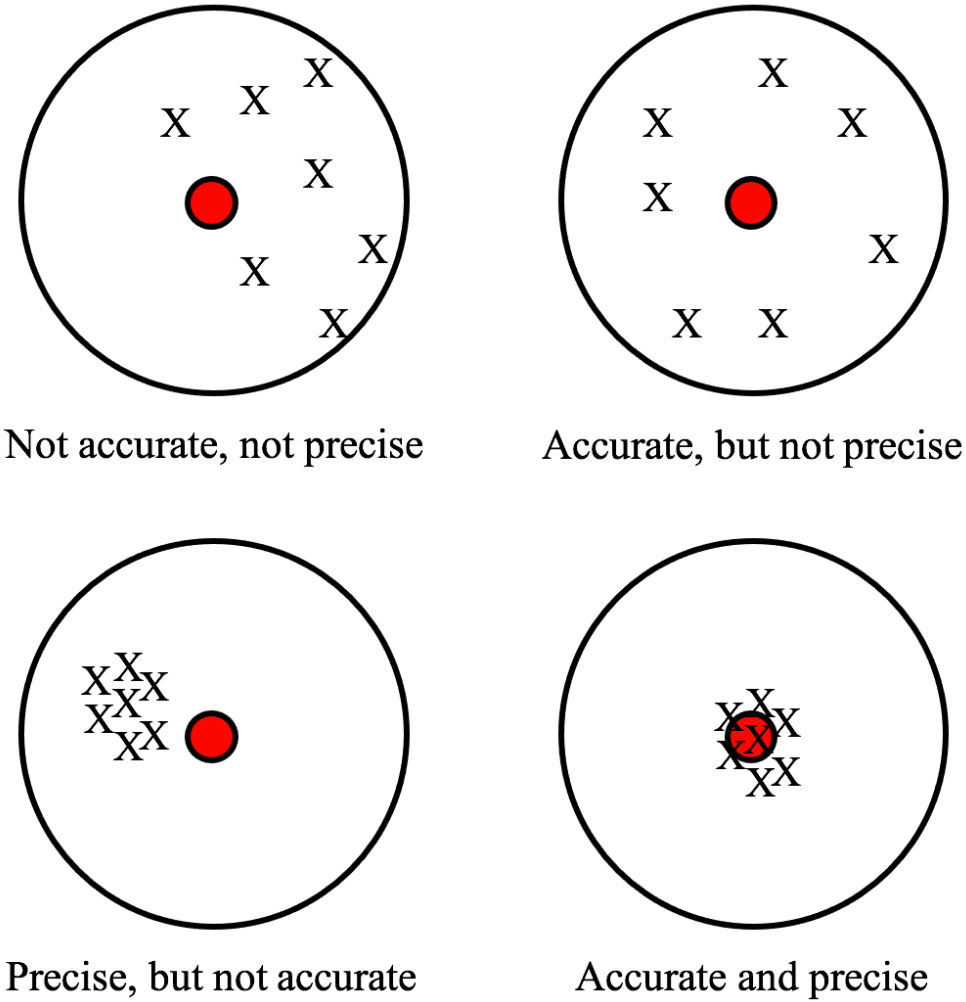
(the assignment of treatments to experimental units)

The functions of randomization:

1. **To neutralize systematic biases**

Proper randomization helps provide valid estimates of experimental error and relative treatment means.

2. **To ensure independence of errors, an assumption of many statistical tests.**



A good experimental design is characterized by the **absence of systematic error**. Experimental units should not differ in any systematic way from one another.

Treatment structure

The type and number of treatment levels are important considerations, particularly when the treatments are quantitative:

1. Separation of levels at equal intervals can facilitate comparisons and interpretation.

2. The number of levels sets the limit on the detectable complexity of the response.

A final aesthetic consideration

A good experimental design is
as simple as possible
for the desired objective.

Relative precision of designs involving few treatments

Precision, sensitivity, or amount of information is measured as the reciprocal of the variance of the means. If we let $I_{\bar{Y}(n)}$ represent the amount of information contained in a sample mean, then:

$$I_{\bar{Y}(n)} = \frac{1}{\sigma^2_{\bar{Y}(n)}} = \frac{n}{\sigma^2}$$

Thus the amount of information *per observation* is:

$$I = \frac{1}{\sigma^2}$$

If s^2 is used to estimate σ^2, there is a correction to this formula:

$$I = \frac{(df+1)}{(df+3)}\frac{1}{s^2}$$

Note that when n → ∞, then the correction factor (df+1)/(df+3) → 1.

The amount of information provided by each experimental unit (i.e. independent observation) in a given experiment is:

$$I = \frac{(df_e + 1)}{(df_e + 3)} \frac{1}{MSE}$$

The *relative efficiency* of Design 1 relative to Design 2 is calculated as the ratio of the amount of information in the two designs:

$$\mathrm{RE}_{1\ to\ 2} = \frac{I_1}{I_2} = \frac{\frac{(df_{e1}+1)}{(df_{e1}+3)MSE_1}}{\frac{(df_{e2}+1)}{(df_{e2}+3)MSE_2}} = \frac{(df_{e1}+1)(dfe_2+3)MSE_2}{(df_{e2}+1)(df_{e1}+3)MSE_1}$$

If this ratio is greater than 1, Design 1 provides more information and is more efficient than Design 2.

SO...

when designing your own experiments,

when reviewing the experiments of others,

when considering whether or not to contribute your efforts to a pre-conceived experiment,

...there are some basic questions you should answer:

FUNDAMENTAL...

1. **What exactly is the question?**

   What falsifiable hypothesis is genuinely put at risk by this experiment?
   Where does this hypothesis lie in the larger logical tree of hypotheses?

2. **What is the intended scope for the conclusions?**

3. **What constitutes a "valid" result?**

NEXT…

1. **Are the methods appropriate to the question and the intended scope of conclusions (i.e. the Objective)?**

   What are the treatments?
   What is the experimental unit?
   What is the unit of observation?
   How were each selected, manipulated, and observed?

2. **Do the data meet the assumptions of the analysis?**

3. **Is there sufficient replication to "falsify" the hypothesis?**

"Get it right or let it alone.
The conclusion you jump to may be your own."

James Thurber
Further Fables for Our Time