Experimental Design
~ Use of treatments and controls ~

NRC 601
Research Concepts in Natural Resources

Department of Natural Resources Conservation
University of Massachusetts Amherst

Fall 2009

Instructor: *Stephen DeStefano*

Why Collect Data?

If we can conduct a *census*, which is . . .

a complete count, there is no need to collect data.

More often, we cannot count or measure every individual in a population, so we . . .

sample the population and collect *data.*

Sampling involves *variation* and *uncertainty* . . .

statistics uses probability theory to address the uncertainty inherent in sample information.

Research design determines the type and quality of the data you collect, and the inferences that can be made.

(Experimental exclosures for ungulate browsing research.)

Some concepts from classical designs developed for agricultural research can be applied to other situations.

Some Terminology

Population: the entire collection of entities or individuals about which we wish to make an inference.

Sample: a subset of the population (we observe the sample but want to know about the population).

Parameter: the true value of some population attribute, which is almost never known.

Statistic: any quantity computed or estimated from the sample.

Some More Terminology

Probability: a set of mathematical tools used to gauge the confidence we have in sample estimates.

Model: some approximation of reality.

Statistical model: a mathematical expression to help us predict a response variable as a function of one or more explanatory variables, based on a set of assumptions.

Data analysis can take 3 approaches

(1) Summarize information, i.e., descriptive statistics

(2) Estimate population parameters using sample data, i.e., parameter estimation.

(3) Test statistical hypotheses, i.e., inferential statistics.

Why Do Experiments?

Suppose a farmer wants to use a new variety of corn:

Two basic questions:

1. *Is there a real difference in yield?*

   New Variety ≠ Standard Variety

2. *What is the size or magnitude and direction of the difference?*

   Compute an estimate and variance

Most experimentation done for 2 reasons:
Hypothesis Testing and Parameter Estimation

Some Important Concepts

Experimental Error

Replication

Randomization

Blocking

Control

Experimental Error

Suppose the farmer does this:

Plants NVcorn patch near the road and
SVcorn patch up the hill . . . *good design?*

How much variation was due to a real difference
vs.
the natural variation in all biological systems?

Random variation among ***experimental units*** is called ***experimental error***.

You must control experimental error.

Replication

- the best procedure for estimating experimental error

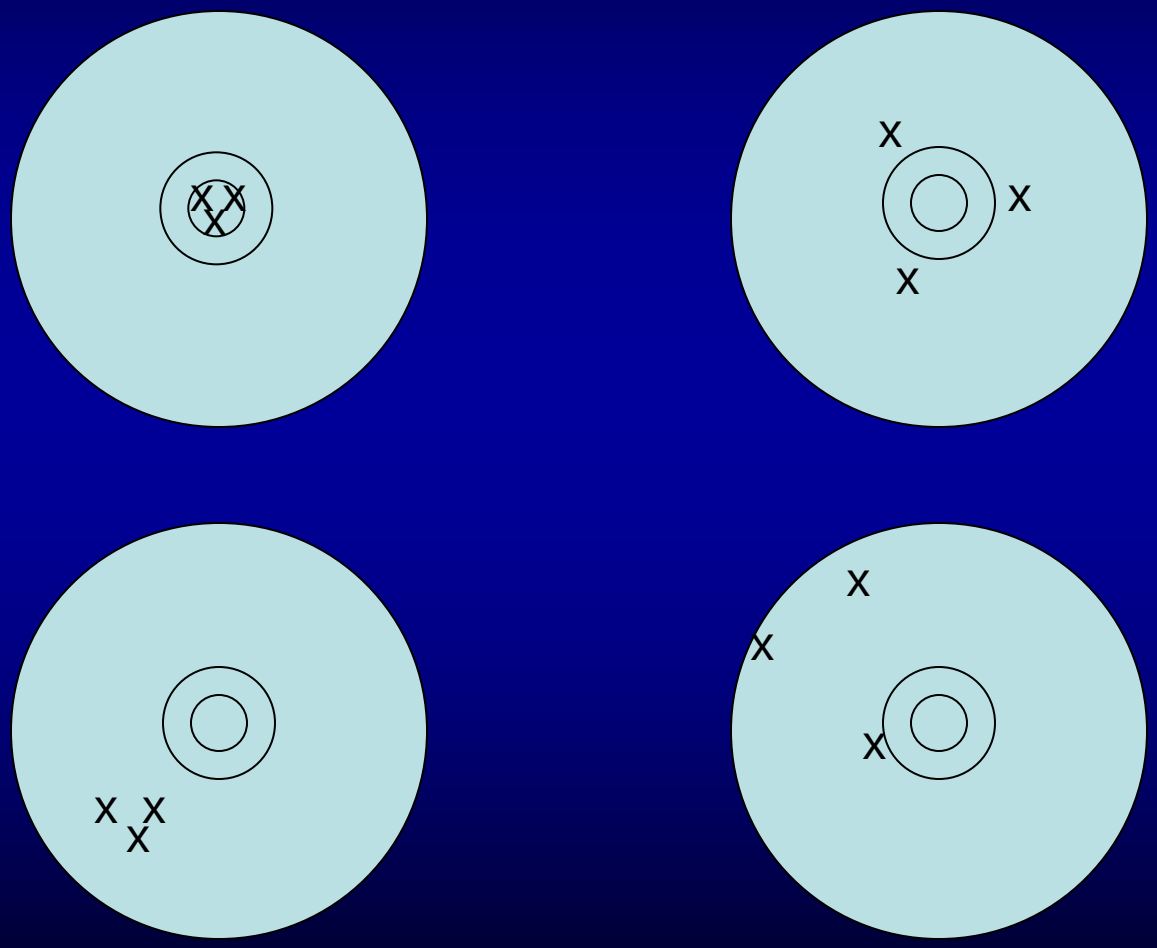
- replication serves a number of purposes:

1. provides an estimate of experimental error (i.e., gives you some idea of natural variation).

2. increases precision by reducing S.E.

3. can broaden the basis for inference because more replicates "capture" a wider variety of conditions.

Accuracy versus Precision



Randomization

- used to assign treatments to E.U.

- no *t* is consistently favored

- no E.U. or subject is consistently favored

- at least 2 good reasons for randomization:

  1. eliminates bias
  2. ensures independence among observations

Blocking

- also called *local control*

- arrange E.U. into groups (blocks) >>> *t* assigned at random to E.U. within the blocks

- a number of good reasons for blocking:

  1. increased precision by removing differences among blocks
  2. treatments are compared under more equal conditions
  3. increased information and scope of inference because of a wider variety of conditions

Controls

- could consider this the "*do nothing*" option . . .

  . . . but is actually another type of "*treatment*"

- absolutely necessary in experimental design

- it is the placebo in medical experiments

- moose exclosures
  - Full exclosure keeps out moose and deer
  - Partial exclosure keeps out moose
  - Controls allow for natural browsing

Completely Randomized Design

1 **D**	*2* **C**	*3* **A**	*4* **A**
5 **C**	*6* **B**	*7* **A**	*8* **C**
9 **B**	*10* **D**	*11* **D**	*12* **B**

(with 4 treatments and 3 replicates)

Randomized Block Design

i	ii	iii	iv
1 **B**	*1* **B**	*1* **D**	*1* **A**
2 **A**	*2* **E**	*2* **E**	*2* **E**
3 **D**	*3* **C**	*3* **A**	*3* **D**
4 **E**	*4* **A**	*4* **C**	*4* **C**
5 **C**	*5* **D**	*5* **B**	*5* **B**

>>>>>> GRADIENT >>>>>>

Latin Square Design

	1	2	3	4
1	A	B	C	D
2	B	C	D	A
3	C	D	A	B
4	D	A	B	C

SLOPE

>>>>>> FERTILITY >>>>>>

Split Plot Design

- assign first level or factor at random to blocks:

  e.g., S_x for seedbed prep. for planting pines

- assign second level to split-plots or subplots

  e.g., N_x for nitrogen fertilizer amount

- in this example, we have 2 blocks (I and II),

  3 whole plots ($S_{1\text{-}3}$), and 4 split- or subplots ($N_{1\text{-}4}$)

Split Plot Design

Block		I	
Plot	S_3	S_1	S_2
Subplot	N_3	N_2	N_0
	N_2	N_3	N_3
	N_1	N_0	N_2
	N_0	N_1	N_1

	II	
S_1	S_3	S_2
N_3	N_0	N_1
N_2	N_1	N_0
N_0	N_3	N_3
N_1	N_2	N_2