STATISTICS 512: DESIGN OF EXPERIMENTS

- **Instructor:** Max Morris

  Phone 515-294-2775, FAX 515-294-5040, mmorris@iastate.edu

  Office hours: MW 11:00 - 11:50, T 10:30 - 11:20, 2103 Snedecor

- **Prerequisite:** STAT 500 & 510 (or old 511) – firm requirement

- **Textbook:** *Design of Experiments: An Introduction Based on Linear Models*, Max Morris, 2010, Chapman & Hall, ISBN 978-1-58488-923-6.

- **Web Page:** http://mmorris.public.iastate.edu/stat512/

- **Course Requirements:**

  1. *Homework:* Semi-regular assignments will be made during the semester. Papers will generally be collected in class one week following the day on which the assignment is made. Some assignments will involve the use of statistical or other software.

2. *Quizzes:* Two closed-book in-class quizzes will be given during the semester ... plan on October 3 and November 7.
3. *Final Exam:* A comprehensive closed-book exam will be given Wednesday, December 17, 9:45 - 11:45 am (tentative University schedule).

- **Grading System:**

Homework	15%	(of course grade)
Quizzes	$2 \times 25\%$	
Final	35%	

Three experiments:

Strength of metal bars:

1. Standard metal rods are processed with one of 3 heat treatments.
2. Each rod is broken, and the amount of force required to accomplish this is recorded.
3. The goal is to determine which of the 3 heat treatments maximizes the breaking force.
4. All testing is done using the same apparatus in the same lab, with one batch of bars.

Blood iron levels in mice:

1. Juvenile mice of a certain strain are fed one of 5 diets from birth.
2. After 6 months, a sample of blood is drawn from each mouse and the iron concentration measured.
3. The goal is to estimate the differences in blood iron levels due to diet.
4. All testing is done in the same lab, during the same time period, with one batch of mice.

Effectiveness of on-line education:

1. Students are enrolled in one of 2 on-line courses that cover the same material.
2. At the end of the course, each student takes (the same) final exam.
3. The goal is to determine any difference in student performance attributable to the course taken.
4. All students are junior engineering majors at the same university.

Experiments are controlled: (point 4 in each example)

- Experiments are conducted under tightly controlled and uniform conditions.
- This minimizes "noise", so that
  - the power of tests is greater
  - the precision of estimates is greater

  for a given sample size.
- But it means that the individual responses are not typical of "reality", e.g. the average 6-month blood iron concentration for mice receiving diet #2 estimates the (population) average for
  - mice fed diet #2, from this batch, housed in this lab, during this period of time, handled by these techicians, ...
- So another investigator working with the same strain and diets
  - in another lab, next year, ...

  shouldn't necessarily expect similar data values for diet #2.

So, experiments are comparative: (point 3)

- Conceptually, say that for experiment 1, every data value coming from diet #2 is influenced by both the experiment and the diet:

$$E(y_{12}) = E_1 + D_2, \quad Var(y_{12}) = \sigma^2$$

- A different experiment 2, using the same diets, might produce diet #2 data:

$$E(y_{22}) = E_2 + D_2, \quad Var(y_{22}) = \sigma^2$$

(if the degree of experimental control is similar)

... not directly comparable.

- But in each experiment, the observed DIFFERENCE between diet #1 and diet #2 has expectation:

$$E(y_{11} - y_{12}) = (E_1 + D_1) - (E_1 + D_2) = D_1 - D_2$$
$$E(y_{21} - y_{22}) = (E_2 + D_1) - (E_2 + D_2) = D_1 - D_2$$

that is, the influence of the particular experiment "cancels out".

- So, even though individual data values and averages from different experiments may not be comparable, *contrasts* are, hence:
  - "to determine which of the 3 heat treatments maximizes the required force"
  - "to estimate the differences in blood iron levels due to diet"
  - "to determine any difference in student performance attributable to course"

  are reasonable experimental goals.
- This is related to the idea of an "experimental control," one exprimental condition that may not be particularly interesting, but serves as a comparison point for more interesting conditions.

Some fundamental elements of experiments:

- **Response:** the quantity to be "observed" (data to be collected) in an experiment (point 2):
  - breaking force (ft-lbs)
  - blood concentration (μg/dL)
  - final exam score (%)
- **Treatment:** the condition (of INTEREST) under which a response is obtained (last part of point 1):
  - 3 heat treatments
  - 5 diets
  - 2 on-line courses

- **Unit:** resources, often physical, OTHER THAN THE TREATMENT needed to obtain a response (first part of point 1):
  - a metal bar
  - a mouse
  - a student

  and even though experimental CONTROL suggests that the units should be as much alike as possible, they can't be identical, and they may (and usual do) influence the response.
- Much attention is given to the way in which available units are *assigned to* treatments, or treatments ar *applied to* units, e.g.
  - a metal bar is heat-treated ...
  - a mouse is fed ...
  - a student is taught ...

Units influence response:

- For experiment 1 and diet #2 using mice 1, 2, 3

$$E(y_{12i}) = E_1 + D_2 + M_i, \quad Var(y_{12i}) = \sigma^2, \quad i = 1, 2, 3$$

- This confuses comparisons since M's
  - aren't known, and
  - can't be assumed to be the same for any two responses
- The solution is to RANDOMIZE units to treatments:
  - with 15 mice in the batch, randomly "assign" 3 to each diet:

$$E(\bar{y}_{12.}) = E_1 + D_2 + E(M), \quad Var(\bar{y}_{12.}) = \tfrac{1}{3}(\sigma^2 + \text{ something})$$
$$E(\bar{y}_{13.}) = E_1 + D_3 + E(M), \quad Var(\bar{y}_{13.}) = \tfrac{1}{3}(\sigma^2 + \text{ something})$$

    e.g. convert unknown fixed unit effects to random effects.

Another argument for randomization:

- Causation versus Association
- Mouse experiment: Suppose
  - you "reach in and grab" one mouse at a time,
  - assign the first 15 to treatment 1, second 15 to treatment 2, ...
  - but unknown to you, some mice have a "fast mouse gene" that helps them delay being selected, and so affects the treatment received, but also affects blood chemistry ...
- Randomization removes any systematic "common cause" affecting both treatment and response.

Homogeneous "groups" of units:

- bars made from the same batch of material
- mice from the same litter
- students with similar GPA's

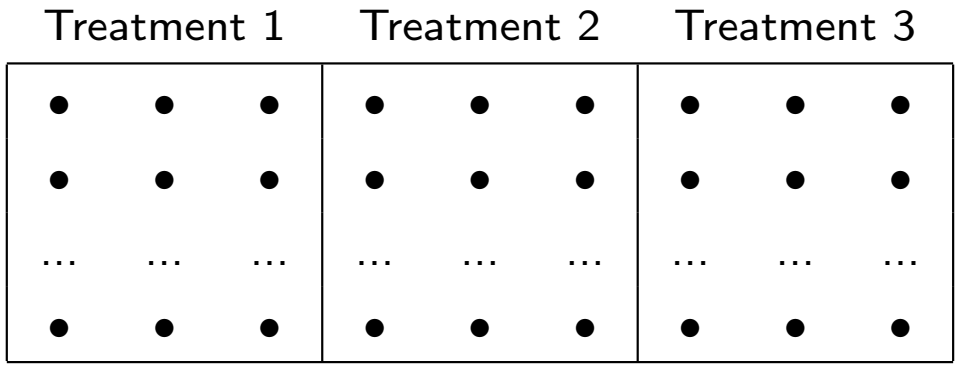
It makes no sense to randomize these across groups; this would increase the "something" component of noise by more than is necessary. *Blocking* refers to *restricted randomization* within homogeneous groups of units:

- with 3 litters of 5 mice each,
- within each litter, randomly assign 1 mouse to each diet

Blocks are then parallel "sub-experiments":

- effects common to all units in a block "cancel out" in treatment comparisons within blocks
- "something" reflects only variation among units *within* blocks

Replication: The evaluation of each treatment more than once – reduces variability/improves properties of inference associated with treatment effects.



- *Differences due to treatments* are estimated by comparing units receiving different treatments
- *Differences due to noise* are assessed by comparing units receiving the same treatment
- Need information about both to decide whether apparent treatment differences are real
- So it is important that the relationship between at least some same-treatment units be the same as between different-treatment units. (When this isn't done, additional assumptions must be made.)

In **STAT 512**, we'll talk about experimental designs with:

- Single and multiple (simultaneous) systems of blocking
- Unstructured and factorial treatment structures
- How split-plot experiments work (where physical entities are units for some purposes and blocks for others)
- How the combined size of the experiment (number of units), treatment assignments, and blocking structure affect statistical inference concerning treatments:
  - estimability of treatment contrasts
  - precision of estimates
  - power of tests