The Design of Statistical Studies (Ch 1-2)

Yifan Zhu

Iowa State University

Outline

Handling Extraneous Variables

Extraneous variables: variables that could influence the response but which are not of practical interest

Ignoring the extraneous variables in the experiment planning can cloud the perception of the effect of treatment variables that are of interest - need to assign treatments to experimental units in a way that remove the effect of extraneous variables

Handling Extraneous Variables

Two ways to handle extraneous variables:

- **blocking:** include the variables as an experimental factor, with the purpose of creating relatively homogeneous environments in which to look for the effect of the treatment variables - effect of the blocking variable is removed within each block
- **randomization:** not all extraneous variables can be supervised - using a randomizing device or table of random digits in choice of experimental protocal for each experimental unit - hope is to balance out the effect of extraneous variables

Outline

{ no blocking only randomization

→ blocking & randomization

Completely Randomized Design

- **Completely Randomized Design**
  - an experimental design with one treatment variable and no blocking variables.
  - Sample units are randomly assigned to treatment levels.
- Example: metallurgy
  - Test the effect of different additives on the corrosion rate of steel.
  - Sample: 12 pieces of raw iron
  - Treatment: additive (A, B, or C).
  - Treatment groups: A (units 1-4), B (units 5-8), and C (units 9-12)

treatments

3 treatment groups

result of randomization

Sample unit	Additive	Sample unit	Additive
1	A	7	B
2	A	8	B
3	A	9	C
4	A	10	C
5	B	11	C
6	B	12	C

numbers are assigned after randomization

Factorial Design

- **Factorial Design**
  - an experimental design with multiple treatment variable (as factors) and no blocking variables.
  - Each sample unit is randomly assigned to a combination of treatment levels.
- Example: metallurgy: a 3×2 factorial version
  - Treatment 1: additive (A, B, or C).
  - Treatment 2: temperature (high or low)
  - Treatment groups: A high (units 1-2), A low (units 3-4), B high (units 5-6), B low (units 7-8), C high (units 9-10), C low (units 11-12),

Unit	Additive	Temp	Unit	Additive	Temp
1	A	high	7	B	low
2	A	high	8	B	low
3	A	low	9	C	high
4	A	low	10	C	high
5	B	high	11	C	low
6	B	high	12	C	low

Metallurgy example: a 2^3 factorial version

- Sample: 16 pieces of iron.
- Treatments:
  - Treatment 1: additive (A or B)
  - Treatment 2: temperature (high or low)
  - Treatment 3: smelting time (long or short)
- Treatment groups: A high long (units 1-2) A high short (units 3-4), ..., B low short (units 11-12).

Unit	Add	Temp	Smelt
1	A	high	long
2	A	high	long
3	A	high	short
4	A	high	short
5	A	low	long
6	A	low	long
7	A	low	short
8	A	low	short

Unit	Add	Temp	Smelt
9	B	high	long
10	B	high	long
11	B	high	short
12	B	high	short
13	B	low	long
14	B	low	long
15	B	low	short
16	B	low	short

Treat 1 : additive (A or B) 2 level

Treat 2 : temp (high or low) 2 level

Treat 3 : Smelting time (long or short) 2 level

Treat 4 : machine (A, B, C) 3 level

$2 \times 2 \times 2 \times 3 = 24$ combos

A, high, long, A A, high, long, B .. .. .., C	A, low, long, A A, low, long, B .. .. .., C
A, high, short, A .. .. .. B C	A, low, short, A .. .. .. B C

→ A → B

Randomized Complete Block Design

- **Randomized Complete Block Design**
  - an experimental design with one or more treatment variable and at least one blocking variable.
  - Within each block separately, sample units are assigned to treatment groups
- Example: metallurgy
  - Treatment: additive (A, B, or C).
  - Blocking variable: pig iron supplier (Amset or Miller and Co.)

6 units 6 units

Unit	Supplier	Add
1	Amset	A
2	Amset	A
3	Amset	B
4	Amset	B
5	Amset	C
6	Amset	C

Unit	Supplier	Add
7	Miller	A
8	Miller	A
9	Miller	B
10	Miller	B
11	Miller	C
12	Miller	C

Outline

Simple Random Sampling

- **Simple Random Sampling**: drawing a sample of n units from a finite population of N units such that every possible n-sized subset of the population has an equal chance of being selected.
- Use either a computerized random number generator or a table of random digits.

Random Digits

12159	66144	05091	13446	45653	13684	66024	91410	51351	22772
30156	90519	95785	47544	66735	35754	11088	67310	19720	08379
59069	01722	53338	41942	65118	71236	01932	70343	25812	62275
54107	58081	82470	59407	13475	95872	16268	78436	39251	64247
99681	81295	06315	28212	45029	57701	96327	85436	33614	29070

Steps of Simple Random Sampling

1. Let M be the number of digits in the number $N - 1$, where N is the population size. (If $N = 1000$ then $M = 3$ digits.)
2. Give each member of the population an M-digit index, i (say, $i = 000, 001, \ldots, 999$)
3. Move through the table of random digits from left to right, top to bottom, selecting population members for the sample when you encounter their indices (ignoring indices that have already been chosen) until you have selected n units for the sample.

Random Digits

12159	66144	05091	13446	45653	13684	66024	91410	51351	22772
30156	90519	95785	47544	66735	35754	11088	67310	19720	08379
59069	01722	53338	41942	65118	71236	01932	70343	25812	62275
54107	58081	82470	59407	13475	95872	16268	78436	39251	64247
99681	81295	06315	28212	45029	57701	96327	85436	33614	29070

Your turn: metallurgy

Using the table of random digits below, take a simple random sample of 12 units of pig iron out of a shipment of 90 units.

$N = 90$, $M = 2$

12159	66144	05091	13446	45653
30156	90519	95785	47544	66735
59069	01722	53338	41942	65118
54107	58081	82470	59407	13475
99681	81295	06315	28212	45029

1. label. 1, 2, ... 90. $M = 2$

12, 15, 61, 44, 5, 9, 11, 34, 46, 45, 65, 33. → 12 sample units

Your turn: metallurgy

Solution:

- Indexed the members of the population from 00 to 89.
- Selected units 12, 15, 61, 44, 5, 9, 11, 34, 46, 45, 65, and 33 for the sample.

12159	66144	05091	13446	45653
30156	90519	95785	47544	66735
59069	01722	53338	41942	65118
54107	58081	82470	59407	13475
99681	81295	06315	28212	45029

Outline

Randomization

- **Randomization**: assigning sample units to treatment groups in an experiment such that every set of assignments is equally likely.
- Steps to randomize n sample units to t treatment groups, each of size s ($n = ts$):
  1. Use the table of random digits to select s units for treatment group 1 from the experimental sample of n units.
  2. Continuing from your last spot in the table, select s units for treatment group 2 from the remaining $n - s$ units in the experimental sample.
  3. Continue this process until you have selected $t - 1$ treatment groups. The remaining units will belong to the last treatment group.

Your turn: metallurgy

$n = 12, t = 3, s = 4$

$12 = 3 \times 4$

Randomize our experimental sample of 12 units of pig iron to thee treatment groups (for additives A, B, and C).

Label : 1, 2, 3, ..., 12. $n = 2$

12159	66144	05091	13446	45653
30156	90519	95785	47544	66735
59069	01722	53338	41942	65178
54107	58081	82470	59407	13475
99681	81295	06315	28212	45029

A: 12, 5, 9, 11. C: 2, 3, 4, 10.

B: 1, 6, 7, 8.

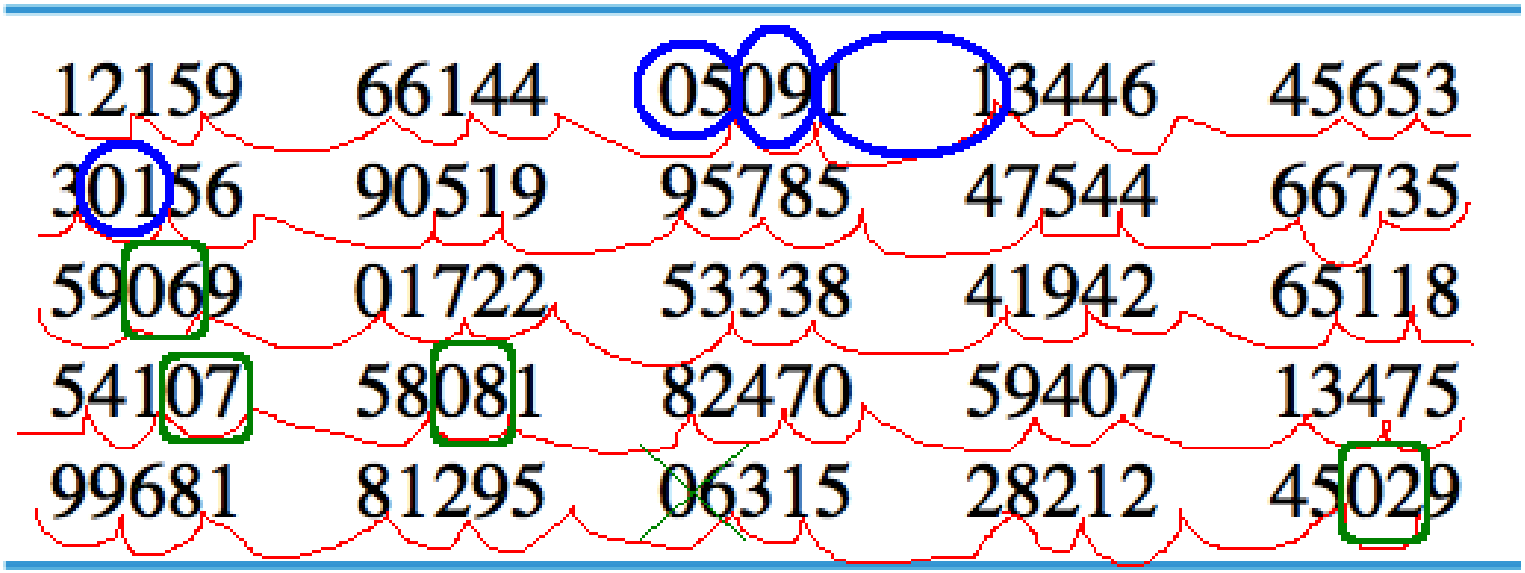
Your turn: metallurgy

labeling: 0, ..., 11.

Solution:

- Units 05, 09, 11, and 01 for group A (blue).
- Units 06, 07, 08, and 02 for group B (green).
- Units 03, 04, 10, and 00 for group C (leftover).

(These unit indices are used for randomization, different from those in Page 6 which are used to index units in different treatment groups)



For randomization in factorial studies, know your treatment groups.

- Example: metallurgy: a 3×2 factorial version
  - Sample: $n = 12$ units
  - Treatment 1: additive (A, B, or C).
  - Treatment 2: temperature (high or low).

6 combos. $t = 6$, $S = 12/6 = 2$

1. How many treatment groups do we have?
2. How many units of the experimental sample should I randomize to each treatment group?

Know you treatment groups: answers

1. $3 \times 2 = 6$ treatment groups.
2. Each treatment group has $12/6 = 2$ units of pig iron.

Ex: For this 3×2 factorial study, we label the six experimental runs as

1: A high, 2: A low

3: B high, 4: B low

5: C high, 6: C low.

Based on the following random digits

46582 18378 63920 39544 25804

What is the order of the experiments?

M = 1.

4, 6, 5, 2, 1, 3

Randomization with blocks

- Randomize units to treatments *within each block*.

Unit	Supplier	Add
1	Amset	A
2	Amset	A
3	Amset	B
4	Amset	B
5	Amset	C
6	Amset	C

Unit	Supplier	Add
7	Miller	A
8	Miller	A
9	Miller	B
10	Miller	B
11	Miller	C
12	Miller	C

- For the metallurgy block design:
  - Randomize all *Amset units* to treatments A, B, and C
  - Then, picking up where you left off in the table of random digits, randomize all *Miller units* to treatments A, B, and C.

Your turn: metallurgy block design

labeling: 1, 2, ... 6.

- Given:
  - 2 blocks (Amset and Miller).
  - 3 treatment levels (A, B, and C).
- Randomize the 12 units of pig iron to treatment groups

12139	66144	05091	13446	45653
30156	90519	95785	47544	66735
59069	01722	53338	41942	65118
54107	58081	82470	59407	13475
99681	81295	06315	28212	45029

A: 1, 2. C: 3, 4.
B: 5, 6.

The Amset Block

- Index the 6 Amset units of pig iron from 0 to 5.
- Using the table of random digits, select:
  - Units 1 and 2 for group A (blue).
  - Units 5 and 4 for group B (green).
  - Units 0 and 3 for group C (leftover).

12159	66144	05091	13446	45653
30156	90519	95785	47544	66735
59069	01722	53338	41942	65118
54107	58081	82470	59407	13475
99681	81295	06315	28212	45029

The Miller Block

- Index the 6 Miller units of pig iron from 0 to 5.
- Using the table of random digits, select:
  - Units 4 and 0 for group A (orange).
  - Units 5 and 1 for group B (red).
  - Units 2 and 3 for group C (leftover).

Continue from the last digit for blocks

12159	66144	05091	13446	45653
30156	90519	95785	47544	66735
59069	01722	53338	41942	65118
54107	58081	82470	59407	13475
99681	81295	06315	28212	45029