STAT 140: Design of Experiments

Section 1: Foundations and Assignment Mechanisms

February 1, 2026

Overview

This worksheet covers fundamental concepts in experimental design including basic definitions, R programming basics, potential outcomes, SUTVA, Fisher's three principles, and assignment mechanisms.

1 Key Definitions

- **Experiment**: A scientific procedure undertaken to make a discovery, test a hypothesis, or demonstrate a known fact.
- **Response**: The measured outcome of the experiment.
- **Experimental factors**: Variables that can have different levels. Often, the active level (e.g., administering medicine) is the **treatment**, while the passive level (e.g., placebo) is the **control**.
- **Experimental units**: The physical entities which can be assigned, at random, to a treatment.
- **Estimands**: The quantities the experimenters aim to estimate from the experiment.
- **Potential outcomes**: The outcome that would be realized if an individual received a specific treatment value: $Y_i(0)$, $Y_i(1)$.

Running Example: The Soy Sauce Experiment

Mr. Paul investigates two brands of soy sauce (existing vs. new) with 50 statistics department members. Each person rates fried rice on a 1-5 scale. The goal: estimate the average causal effect of the new soy brand.

- **Units**: 50 individuals
- **Treatment**: New soy sauce ($W_i = 1$)
- **Control**: Existing soy sauce ($W_i = 0$)
- **Response**: Taste rating (1-5)
- **Estimand**: Average treatment effect

2 Review: Combinations and Factorials

Factorials

$$n! = n \times (n-1) \times (n-2) \times \ldots \times 2 \times 1$$

Combinations

The number of ways to choose r items from n when order doesn't matter:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Exercise: Treatment Assignment Counting

(i) List all possible ways to assign 2 units (out of 5) to treatment.

(ii) Use the combinations formula to calculate the number of assignment vectors.

3 Potential Outcomes Framework

Notation

- $Y_i(0)$: Outcome for unit i under control
- $Y_i(1)$: Outcome for unit i under treatment
- W_i: Treatment assignment (0 or 1)
- Y_i^{obs}: Observed outcome
- Y_i^{mis}: Missing (unobserved) outcome

Observed vs. Missing Outcomes

$$Y_i^{\text{obs}} = \begin{cases} Y_i(0) & \text{if } W_i = 0 \\ Y_i(1) & \text{if } W_i = 1 \end{cases} = W_i \cdot Y_i(1) + (1 - W_i) \cdot Y_i(0)$$

$$Y_i^{\text{mis}} = \begin{cases} Y_i(1) & \text{if } W_i = 0 \\ Y_i(0) & \text{if } W_i = 1 \end{cases}$$

The Science Table

Unit i	$Y_i(0)$	$Y_i(1)$	$\tau_i = Y_i(1) - Y_i(0)$
1	?	?	?
2	?	?	?
$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	?	?	?

The Fundamental Problem: We can NEVER observe both $Y_i(0)$ and $Y_i(1)$ for the same unit

Exercise 1.2: The Science Table

For Mr. Paul's experiment with 8 members, create the science table. What can we observe if units 1, 3, 5, 7 receive the new soy sauce?

4 SUTVA: Stable Unit Treatment Value Assumption

The potential outcomes are a function of the unit and treatment only.

1. **No interference between units:** The potential outcomes for unit i do not vary with treatments assigned to other units.
2. **No hidden variations of treatments:** For each unit, there are no different versions of each treatment level that lead to different potential outcomes.

Exercise: SUTVA Violations

For each scenario, identify if SUTVA is violated and explain why:

(a) In a vaccine trial, vaccinated individuals reduce disease transmission to unvaccinated individuals.

(b) Testing two ”identical” drug pills, but one was stored properly and one was exposed to heat.

(c) In the soy sauce experiment, participants can smell each other’s food.

5 Fisher’s Three Fundamental Principles

5.1 Randomization

Why randomize?

(i) **Prevents confounding:** Treatment-control comparison is not confounded with unobserved covariates

(ii) **Reduces subjective bias:** Prevents conscious/unconscious bias in assignment

(iii) **Ensures valid inference:** Creates known probability distribution for hypothesis tests

> Randomization is what allows us to make **causal claims** rather than just observational associations.

5.2 Blocking

- Divide experimental units into homogeneous groups (blocks)
- Randomly assign treatments within each block
- Reduces variability, increases precision
- Example: Block by spice preference (mild vs. spicy lovers), then randomize soy sauce within each block

5.3 Replication

- Assign multiple units to each treatment
- **Why needed?**
  - Estimate variability
  - Distinguish treatment effects from random noise
  - Enable valid statistical conclusions

6 Assignment Mechanisms

The **assignment mechanism** is the process that determines which units receive which treatments, hence which potential outcomes are observed and which are missing.
Formally: A function $P(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$ that assigns probabilities to all 2^N possible assignment vectors.

7 Three Common Assignment Mechanisms

7.1 Completely Randomized Design (CRD)

Procedure: Fix N_T and N_C in advance. Randomly choose which N_T units receive treatment.
Assignment Mechanism:

$$P(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \begin{cases} \binom{N}{N_T}^{-1} & \text{if } \sum_{i=1}^{N} W_i = N_T \\ 0 & \text{otherwise} \end{cases}$$

Key Properties:

- All $\binom{N}{N_T}$ possible assignments are equally likely
- Each unit has probability $p_i = N_T/N$ of treatment
- Sample sizes in each group are fixed

7.2 Bernoulli Assignment Mechanism

Procedure: Flip an independent (possibly biased) coin for each unit.
Assignment Mechanism:

$$P(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \prod_{i=1}^{N} p_i^{W_i}(1 - p_i)^{1-W_i}$$

Often $p_i = 0.5$ for all i (fair coin).
Key Properties:

- Can assign as units arrive sequentially
- Number treated is random: $N_T \sim \text{Binomial}(N, p)$
- All 2^N assignment vectors are possible
- Could get all treated or all control (low probability)

7.3 Paired Randomized Experiment

Procedure: Pair up similar units (based on covariates). Within each pair, randomly assign one to treatment, one to control.
Requirements:

- N units, with N even
- Units grouped into $N/2$ pairs
- Within each pair: one unit gets treatment

Assignment Mechanism:

$$P(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \begin{cases} 2^{-N/2} & \text{if } \mathbf{W} \in \mathcal{W}^+ \\ 0 & \text{otherwise} \end{cases}$$

where $\mathcal{W}^+$ is the set of valid paired assignments.

Comparing Three Assignment Mechanism

Design	**Pros**	**Cons**	**When to Use**
CRD	Fixed sample sizes; efficient	Need to know N_T, N_C in advance	Standard experiments
Bernoulli	Simple; sequential assignment	Random group sizes; extreme outcomes possible	Online experiments
Paired	High precision; controls covariates	Requires pairable units	Twins, matched patients