STAT 140: Design of Experiments

Section 2 Fisherian Inference

February 9, 2026

By the end of this tutorial, you should understand:

- The Fundamental Problem of Causal Inference
- What makes a null hypothesis sharp
- How Fisher Randomization Tests work
- Why randomization → not distributional assumptions → drives inference

1 The Fundamental Problem of Causal Inference

For each individual i, there exist two potential outcomes:

$$Y_i(1) \quad \text{(if treated)}, \qquad Y_i(0) \quad \text{(if not treated/control/placebo)}$$

However, we only observe one:

$$Y_i^{obs} = Y_i(W_i)$$

The missing outcome is called the **counterfactual**.

We can never observe both potential outcomes for the same individual. This is the Fundamental Problem of Causal Inference.

2 Sharp Null Hypothesis

A null hypothesis is called **sharp** if it specifies the exact value of every missing potential outcome.

$$H_0 : Y_i(1) = Y_i(0) \quad \forall i$$

In plain English, this means there is no treatment effect.

A sharp null removes the counterfactual problem because all missing outcomes can be filled in exactly.

This allows us to simulate outcomes under every possible treatment assignment:

- No modelling required
- No normality assumptions
- Inference comes purely from the experimental design

3 Choosing a Test Statistic

A test statistic summarizes evidence against the null.
Common choice:

$$T(W, Y^{obs}) = \frac{1}{N_1} \sum_{i:W_i=1} Y_i^{obs} - \frac{1}{N_0} \sum_{i:W_i=0} Y_i^{obs}$$

Large values of $|T|$ indicate stronger evidence of a treatment effect.

4 Worked Example: Training Program and Test Scores

Suppose four students participate in a training program designed to improve exam performance.

Since the program is intended to help students, we are primarily interested in detecting a positive treatment effect.

Sharp Null Hypothesis:

$$H_0 : Y_i(1) = Y_i(0) \quad \forall i$$

(Treatment has no effect)

One-Sided Alternative:

$$H_A : Y_i(1) > Y_i(0)$$

(The training program improves scores)

The observed scores are:

Student	Score
A	90
B	80
C	70
D	60

Observed Assignment:

Treated	Control
90	70
80	60

We use the difference in means as our test statistic:

$$T = \text{mean(treated)} - \text{mean(control)}$$

$$T^{obs} = 85 - 65 = 20$$

A large positive value seems to provide evidence that the training program is effective.

5 Fisher Null Distribution

Under the sharp null, outcomes remain fixed regardless of treatment assignment.

There are:

$$\binom{4}{2} = 6$$

possible assignments.

Treated Pair	Statistic T
(90, 80)	20
(90, 70)	10
(90, 60)	0
(80, 70)	0
(80, 60)	-10
(70, 60)	-20

6 Compute the One-Sided p-value

Since our alternative hypothesis is that the training program increases scores, we only consider statistics at least as large as the observed value.

$$T \geq 20$$

Only one assignment satisfies this condition.

$$p = \frac{1}{6} \approx 0.17$$

Even though the treated group scored higher, such a difference could still reasonably occur due to random assignment alone. We therefore fail to reject the sharp null hypothesis.

7 Practice Problems

7.1 Problem 1: The Honey Experiment

An experiment was done to evaluate the effect of honey treatment on nocturnal cough frequency in children (see Chapter 5 of the Readings). The outcome is measured on a scale from zero ("not at all frequent/severe") to six ("extremely frequent/severe"). Let us consider, for relative ease of exposition, a sub-sample of the honey data set, with six children. The subsequent table gives the observed data on cough frequency for these six children in the potential outcome form.

W_i^{obs}	$Y_i(0)$	$Y_i(1)$
1		3
1		5
1		0
0	4	
0	0	
0	1	

We are interested in estimating the causal effect of buckwheat honey on cough.

1. Specify Fisher's sharp null hypothesis of no treatment effect in mathematical notation.

2. Under Fisher's sharp null hypothesis, fill in all six of the missing potential outcomes in the table.

3. We use the difference-in-means statistic $T = \bar{Y}^{obs}(1) - \bar{Y}^{obs}(0)$. Compute the observed value of this test statistic.

4. In a completely randomized experiment with 6 children, 3 are assigned to treatment and three to control. How many distinct treatment assignment vectors are possible? Show your calculation.

5. The RMarkdown file provided for this section contains code to simulate the randomization distribution of the test statistic under the stated Fisher's sharp null hypothesis. Using this code:

   (a) Plot a histogram of the randomization distribution of the test statistic.

   (b) Add a vertical line indicating the observed test statistic.

   (c) Compute the p-value using the randomization distribution.

   (d) State your statistical conclusion regarding Fisher's sharp null hypothesis with the p-value computed.

7.2 Problem 2: The Private Tutoring Experiment

An experiment was designed to study the effect of private tutoring ($W_i = 1$) on students' test scores. We are in a completely randomized experiment setting and the observed data are

$$W^{obs} = (0, 0, 0, 0, 0, 1, 1, 1, 1, 1), \qquad Y^{obs} = (2, 9, 2, 10, 3, 7, 5, 8, 9, 11).$$

We wish to assess the hypothesis that for all units the treatment increases test scores by a constant amount $C = 2$:

$$H_{cst} : Y_i(1) = Y_i(0) + 2.$$

Our test statistic is the absolute difference between the treated and control averages minus the hypothesized increase:

$$T = \left| \frac{1}{N_T} \sum Y_i(1) - \frac{1}{N_C} \sum Y_i(0) - C \right|.$$

1. Construct the observed data table in potential outcome form and impute the missing potential outcomes under the hypothesis H_{cst}.

2. Using the imputed potential outcomes, compute the observed value of the test statistic T.

3. Using the RMarkdown file provided for this section:

   (a) Plot the simulated randomization distribution of the test statistic.

   (b) Add a vertical line indicating the observed test statistic.

   (c) Estimate the p-value as the proportion of simulated statistics at least as extreme as the observed value.

   (d) State your statistical conclusion regarding the null hypothesis H_{cst}.

7.3 Problem 3: The Diet Experiment

A scientist wants to test the effect of Diet A (treatment) versus a more traditional Diet B (control). He has 20 animals available for the experiment. The animals differ with respect to age, sex, and other characteristics. The response is the percentage decrease in the animals' weight after two months of treatment. The observed results are shown in Table 1.

Diet A	1	2	3	4	5	6	7	8	9	10
Response	13.2	8.2	10.9	14.3	10.7	6.6	9.5	10.8	8.8	13.3
Diet B	11	12	13	14	15	16	17	18	19	20
Response	14.0	8.8	11.2	14.2	11.8	6.4	9.8	11.3	9.3	13.6

Table 1: Observed results of the diet experiment

Each animal receives either Diet A or Diet B. Allocation was decided by assigning half of the animals to the treatment using a completely randomized design.

1. What is the total number of possible treatment assignments under this completely randomized design?

2. Perform a Fisher randomization test of the sharp null hypothesis of no treatment effect using the difference-in-means statistic.

   (a) Compute the observed difference in means.

   (b) Using **10,000 randomizations**, approximate the p-value.

   (c) Plot a histogram of the randomization distribution and indicate the observed test statistic with a vertical line.

   (d) State your statistical conclusion regarding Fisher's sharp null hypothesis with the p-value computed.

3. What is the smallest possible p-value that could be obtained from this experiment? Briefly explain your reasoning.