Randomized Experiments

Statistics · Module 11 — Causality

Node 11.4

Position in Knowledge Graph

System: Statistics. Structural Domain: Causality (Module 11). Node 11.4.

Randomized experiments are the gold standard for causal identification: they make ignorability hold by design, eliminating confounding without requiring knowledge of the causal structure. This node develops the theory of completely randomized experiments, including the Neyman framework for ATE estimation, Fisher's exact randomization test, and covariate adjustment via ANCOVA. It operationalizes the potential outcomes framework (Node 11.1) under the strongest possible identification conditions, and sets the benchmark against which observational methods (Node 11.5) are evaluated.

Formal Definition

In a **completely randomized experiment (CRE)**, n units are assigned to treatment ($T_i = 1$) or control ($T_i = 0$) by a known random mechanism, with n_1 treated and $n_0 = n - n_1$ control units. The assignment satisfies:

$$\{Y_i(0), Y_i(1)\}_{i=1}^{n} \perp\!\!\!\perp (T_1, \ldots, T_n),$$

and the assignment mechanism is fully specified: $\mathbb{P}(\mathbf{T} = \mathbf{t}) = \binom{n}{n_1}^{-1}$ for all $\mathbf{t}$ with $\sum t_i = n_1$.

Intuition

Randomization physically severs all arrows into the treatment variable in the causal DAG. No matter how complex the confounding structure, random assignment ensures that treated and control groups are comparable on average—both on observed and unobserved characteristics. The beauty of randomization is that identification follows from the design, not from modeling assumptions. The price is feasibility: many causal questions cannot be answered by experiments (ethics, cost, logistics).

Structural Role

Dependencies: Builds on the potential outcomes framework (Node 11.1) for the estimand definitions and hypothesis testing (Node 5.1) for the inferential framework. **Enables:** Observational methods (Node 11.5) by providing the ideal benchmark against which observational designs are compared.

Randomization removes all backdoor paths by design (Node 11.3). Fisher's test provides exact finite-sample inference; Neyman's framework provides asymptotically valid confidence intervals. ANCOVA leverages pre-treatment covariates for precision gain without risking bias.

Mathematical Development

Identification Under Randomization

Under CRE and SUTVA:

$$\mathbb{E}[Y \mid T = 1] = \mathbb{E}[Y(1) \mid T = 1] = \mathbb{E}[Y(1)],$$

where the last equality uses independence. Similarly $\mathbb{E}[Y \mid T = 0] = \mathbb{E}[Y(0)]$. Therefore:

$$\tau_{\text{ATE}} = \mathbb{E}[Y \mid T = 1] - \mathbb{E}[Y \mid T = 0].$$

This requires no modeling, no covariate adjustment, and no distributional assumptions. It follows entirely from the randomization mechanism.

Neyman's Repeated Sampling Framework

The **difference-in-means** estimator is

$$\hat{\tau} = \bar{Y}_1 - \bar{Y}_0 = \frac{1}{n_1} \sum_{i:T_i=1} Y_i - \frac{1}{n_0} \sum_{i:T_i=0} Y_i.$$

Theorem (Neyman, 1923). Under CRE:

$$\mathbb{E}[\hat{\tau}] = \tau_{\text{ATE}}.$$

Proof. $\mathbb{E}[\bar{Y}_1] = \mathbb{E}\Big[\frac{1}{n_1} \sum_{i:T_i=1} Y_i(1)\Big]$. Under random assignment, each unit i has probability n_1/n of being treated, and $\mathbb{E}[\sum_{i:T_i=1} Y_i(1)] = \frac{n_1}{n} \sum_{i=1}^{n} Y_i(1)$, so $\mathbb{E}[\bar{Y}_1] = \frac{1}{n} \sum_i Y_i(1) = \bar{Y}(1)$. Similarly $\mathbb{E}[\bar{Y}_0] = \bar{Y}(0)$. Therefore $\mathbb{E}[\hat{\tau}] = \bar{Y}(1) - \bar{Y}(0) = \tau_{\text{ATE}}$.

Variance of $\hat{\tau}$:

$$\text{Var}(\hat{\tau}) = \frac{S_1^2}{n_1} + \frac{S_0^2}{n_0} - \frac{S_\tau^2}{n},$$

where $S_t^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i(t) - \bar{Y}(t))^2$ and $S_\tau^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\tau_i - \bar{\tau})^2$ is the variance of individual treatment effects.

The term S_τ^2/n involves both potential outcomes and is **not estimable** from data (the fundamental problem). The **conservative variance estimator** is

$$\hat{V} = \frac{s_1^2}{n_1} + \frac{s_0^2}{n_0}, \quad s_t^2 = \frac{1}{n_t - 1} \sum_{i:T_i=t} (Y_i - \bar{Y}_t)^2,$$

which overestimates $\text{Var}(\hat{\tau})$ by $S_\tau^2/n \geq 0$. Under constant treatment effects ($\tau_i = \tau$ for all i), $S_\tau^2 = 0$ and the estimator is exact.

Fisher's Randomization Test

Sharp null hypothesis:

$$H_0^F : Y_i(1) = Y_i(0) \quad \text{for all } i = 1, \ldots, n.$$

Under H_0^F, both potential outcomes equal Y_i^{obs}, so the complete data table is known. For any test statistic $T(\mathbf{T}, \mathbf{Y}^{\text{obs}})$, the exact null distribution is obtained by enumerating all $\binom{n}{n_1}$ possible treatment assignments:

$$p\text{-value} = \frac{|\{\mathbf{t} : T(\mathbf{t}, \mathbf{Y}^{\text{obs}}) \geq T_{\text{obs}}\}|}{\binom{n}{n_1}}.$$

Properties: - Exact for any sample size (no asymptotic approximation). - No distributional assumptions on Y. - Valid for any test statistic (difference in means, rank statistics, Kolmogorov–Smirnov, etc.). - Tests the sharp null (no effect for any unit), which is stronger than Neyman's null ($\tau_{\text{ATE}} = 0$).

Computational note. For large n, exact enumeration is infeasible. Monte Carlo approximation: sample B random permutations and compute the proportion exceeding T_{obs}. This gives a randomization p-value with precision $O(1/\sqrt{B})$.

Covariate Adjustment: ANCOVA

Even under randomization, pre-treatment covariates X can improve precision.

Lin's ANCOVA estimator. Fit the regression:

$$Y_i = \alpha + \tau T_i + \gamma^\top X_i + \delta^\top (T_i \cdot (X_i - \bar{X})) + \varepsilon_i,$$

where the interaction $T_i \cdot (X_i - \bar{X})$ allows different slopes in treated and control groups. The coefficient $\hat{\tau}^{\text{Lin}}$ on T_i is the ANCOVA estimator.

Theorem (Lin, 2013). Under randomization, $\hat{\tau}^{\text{Lin}}$ is:

1. Consistent for τ_{ATE} regardless of whether the linear model is correctly specified.
2. Asymptotically at least as efficient as the unadjusted $\hat{\tau}$.
3. Strictly more efficient when X is predictive of Y.

Why adjustment does not bias. Under randomization, $T \perp\!\!\!\perp X$, so including X as a covariate does not introduce confounding. The regression coefficient on T estimates the same causal effect with or without covariates; the improvement is purely in variance.

Precision gain. If R^2 is the fraction of outcome variance explained by X, the ANCOVA variance is approximately $(1 - R^2) \cdot \text{Var}(\hat{\tau})$. For $R^2 = 0.5$, the effective sample size doubles.

Stratified (Block) Randomization

In **stratified randomization**, units are grouped into strata based on X, and randomization occurs within each stratum. This guarantees exact covariate balance and further reduces variance.

The stratified estimator is

$$\hat{\tau}^{\text{strat}} = \sum_s \frac{n_s}{n} \hat{\tau}_s,$$

where $\hat{\tau}_s = \bar{Y}_{1,s} - \bar{Y}_{0,s}$ is the within-stratum difference in means and n_s is the stratum size. This is guaranteed to have $\text{Var}(\hat{\tau}^{\text{strat}}) \leq \text{Var}(\hat{\tau})$.

Worked Examples

Example 1: Neyman Inference for a Drug Trial

$n = 200$ patients randomized equally: $n_1 = n_0 = 100$. Outcome: blood pressure reduction (mmHg).

Observed: $\bar{Y}_1 = 12.3$, $\bar{Y}_0 = 8.7$, $s_1^2 = 25.0$, $s_0^2 = 20.0$.

$\hat{\tau} = 12.3 - 8.7 = 3.6$ mmHg. $\hat{V} = 25/100 + 20/100 = 0.45$. SE $= \sqrt{0.45} = 0.671$.

95% CI: $3.6 \pm 1.96 \times 0.671 = (2.28, 4.92)$. The CI is conservative (it overestimates the true variance by omitting S_τ^2/n).

Test $H_0 : \tau = 0$: $z = 3.6/0.671 = 5.37$, $p < 0.001$. Strong evidence of a treatment effect.

Example 2: Fisher Randomization Test

$n = 8$, $n_1 = 4$. Observed outcomes: treated $= \{7, 9, 6, 8\}$, control $= \{3, 5, 4, 2\}$.

$T_{\text{obs}} = \bar{Y}_1 - \bar{Y}_0 = 7.5 - 3.5 = 4.0$.

Under H_0^F, all 8 outcomes are fixed. There are $\binom{8}{4} = 70$ possible assignments.

For each permutation, compute $\bar{Y}_1 - \bar{Y}_0$. Count the number of permutations yielding ≥ 4.0.

The maximum possible difference (assigning $\{7, 8, 9, 6\}$ vs. $\{2, 3, 4, 5\}$) is $7.5 - 3.5 = 4.0$, but other permutations achieving this: $\{9, 8, 7, 6\}$ vs. $\{5, 4, 3, 2\}$ gives 4.0. After enumeration, suppose 2 out of 70 permutations yield ≥ 4.0.

p-value $= 2/70 = 0.029$. Reject H_0^F at the 5% level.

Cross-Links

AI. A/B testing in product development is randomized experimentation with the ATE as the estimand. Multi-armed bandit algorithms extend the experimental framework to sequential settings where treatment assignment adapts based on interim results, trading off exploration and exploitation.

Economics. Randomized controlled trials (RCTs) are the basis for the credibility revolution in program evaluation (Angrist, Duflo, Banerjee). Field experiments in development economics randomize at the village or school level (cluster randomization), requiring cluster-robust variance estimators.

Linear Algebra. The design matrix of a CRE has a specific structure: T is a binary vector with exactly n_1 ones. The hat matrix for the unadjusted estimator is $H = \text{diag}(T/n_1 + (1-T)/n_0)$, and the ANCOVA hat matrix is the orthogonal projection onto the column space of $[T, X, T \cdot X]$.

Quantum Mechanics. Bell tests are randomized experiments in physics: measurement settings (treatments) are randomly assigned to entangled particles, and the outcome statistics test whether local hidden variable models (the causal DAG with unmeasured common cause) can explain the observed correlations. Randomization of measurement settings is essential to close the "freedom of choice" loophole.

Structural Compression

Invariant

Randomization renders treatment independent of all potential outcomes; the observed difference in means is an unbiased estimator of the ATE requiring no model for the outcome and no knowledge of the confounding structure.

Mechanism

Physical randomization removes all backdoor paths between T and Y in the causal DAG. The known assignment mechanism enables exact permutation inference (Fisher) and conservative variance estimation (Neyman). Covariate adjustment reduces variance without introducing bias because $T \perp\!\!\!\perp X$ by design.

Cross-System Echo

Fisher's randomization test is the prototype of permutation tests in nonparametric statistics. In AI, A/B testing is the standard experimental paradigm for causal evaluation of algorithmic changes. In economics, RCTs underpin the credibility revolution. In physics, Bell tests use randomized measurement settings to test fundamental causal structure of quantum mechanics.

Exercises

1. Prove that the difference-in-means estimator $\hat{\tau} = \bar{Y}_1 - \bar{Y}_0$ is unbiased for τ_{ATE} under completely randomized assignment. State explicitly where independence of treatment and potential outcomes is used.
2. Derive the Neyman variance formula $\text{Var}(\hat{\tau}) = S_1^2/n_1 + S_0^2/n_0 - S_\tau^2/n$. Show that omitting S_τ^2/n yields a conservative estimator.
3. For $n = 6$, $n_1 = 3$, outcomes $= \{10, 4, 7, 3, 8, 5\}$ (first 3 treated), compute the Fisher exact p-value for the difference-in-means test statistic under the sharp null. Enumerate all $\binom{6}{3} = 20$ permutations.
4. Show that under constant treatment effects ($\tau_i = \tau$ for all i), the Neyman variance formula simplifies to $\text{Var}(\hat{\tau}) = (S_0^2/n)(1/n_1 + 1/n_0)$, and the conservative estimator is exact.
5. Prove Lin's result: under randomization, the ANCOVA estimator with treatment-covariate interactions is consistent for τ_{ATE} even when the linear model is misspecified. (Hint: use $T \perp\!\!\!\perp X$ and the law of iterated expectations.)
6. Compare the variance of the unadjusted estimator, the ANCOVA estimator, and the stratified estimator for an experiment with $n = 100$, two strata of equal size, and $R^2 = 0.4$. Quantify the efficiency gain from each adjustment.
7. Argue, structurally, why randomization is the only design that achieves identification of the ATE without assumptions about the functional form of the outcome model or the set of confounders, and explain why observational methods (Node 11.5) can never fully replicate this property.

Statistics · Module 11 — Causality · Node 11.4
Concepts: hypothesis-testing, random-variables
Prerequisites: 11.1, 5.1
Enables: 11.5

hbar.university — own.your.cognition